

VŠB - Technická univerzita Ostrava
Fakulta elektrotechniky a informatiky
Katedra informatiky

SEO analyzátor
SEO analyzer

2011

Tomáš Gottvald

Zadání diplomové práce

Student:

Bc. Tomáš Gottvald

Studijní program:

N2647 Informační a komunikační technologie

Studijní obor:

2612T025 Informatika a výpočetní technika

Téma:

SEO Analyzátor
SEO Analyzer

Zásady pro vypracování:

Masivní rozvoj internetu s sebou přináší obrovské množství internetových stránek. Ke zkvalitnění určitých prezentací slouží jejich optimalizace. Správná optimalizace má pozitivní vliv na pozici v internetových vyhledávačích. Cílem práce je navrhnout a naimplementovat SEO Robota spolu s hodnotícím algoritmem pro internetové stránky a získané informace prezentovat uživateli.

1. Seznamte se s problematikou SEO.
2. Proveďte analýzu současných, volně dostupných internetových SEO nástrojů a jejich hodnotících parametrů.
3. Navrhněte a naimplementujte vlastního SEO Robota.
4. Navrhněte a naimplementujte vlastní hodnotící algoritmus, zpracovávající hodnoty získané SEO Robotem.
5. Ověřte závislosti parametrů na pozici ve vyhledávačích pomocí statistických metod. Ověření, respektive srovnání proveďte u internetových vyhledávacích portálů google.cz a seznam.cz.
6. Navrhněte a naimplementujte vhodnou internetovou aplikaci prezentující získané hodnoty.

Seznam doporučené odborné literatury:

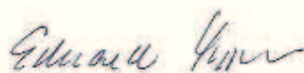
1. Optimalizace pro vyhledávače - SEO, Ing. Radim Smička, ISBN: 80-239-2961-5
2. SEO - Search Engine Optimization, Jennifer Grappone a Gradya Couzin, ISBN: 978-80-86815-85-5
3. 333 tipů a triků pro SEO, Michal Kubiček a Jan Linhart, ISBN: 978-80-251-2468-0

Formální náležitosti a rozsah diplomové práce stanoví pokyny pro vypracování zveřejněné na webových stránkách fakulty.

Vedoucí diplomové práce: **Ing. Petr Madeckí**

Datum zadání: 19.11.2010

Datum odevzdání: 06.05.2011



doc. Dr. Ing. Eduard Sojka
vedoucí katedry



prof. RNDr. Václav Snášel, CSc.
děkan fakulty

Prohlašuji, že jsem tuto diplomovou práci vypracoval samostatně. Uvedl jsem všechny literární
prameny a publikace, ze kterých jsem čerpal.

V Havířově 6. května 2011

.....

Děkuji svým kolegům Bc. Štefanu Bartošovi, Bc. Jiřímu Královi, Bc. Lukáši Heinichovi (abecedně), za podporu a pomoc při testování a za jejich návrhy. Dále pak děkuji za odbornou konzultaci vedoucímu diplomové práce Ing. Petru Madeckému.

Abstrakt

Diplomová práce se zabývá tvorbou SEO analyzátoru ve formě robota, který automaticky v pravidelných cyklech testuje a mapuje webové stránky uživatelů. Výsledky ukládá a v grafické podobě zobrazuje uživateli ve formě grafu. Uživatel tak získává pohodlný nástroj pro sledování testovaných parametrů. Vlastnosti takto vytvořeného robota byly využity ke sledování trendů ve vyhledávačích. Administrátorovi jsou výsledky reprezentovány ve formě grafu nebo ve formě tabulky, kterou lze použít při statistických analýzách. Byla provedena statistická analýza závislosti testovaných parametrů na pozici ve vyhledávačích. Ani u jednoho parametru se však nepodařilo tuto závislost jednoznačně potvrdit. Plnohodnotnou aplikaci však lze použít i pro další zkoumání.

Klíčová slova

SEO, PHP, MySQL, CSS, AJAX, statistika, SEO robot, SEO parametry

Abstract

This thesis deals with creating SEO analyzer in the form of a robot which is automatically testing and mapping user's web site in regular cycles. It saves results and provides them to users in graphic form by the graphs. The user gets a convenient tool to monitor the test parameters. Features such created the robot were used to monitor trends in search engines. Results are represented to Administrator as a graph or table form, which can be used in statistical analysis. Statistical analysis was performed according to the position of the tested parameters in the search engines. This dependence could not be confirmed in any parameter. The full application can be used for further investigation.

Keywords

SEO, PHP, MySQL, CSS, AJAX, statistics, SEO robot, SEO parameters

Seznam použitých zkratk a symbolů

PHP	-	Hypertext preprocessor
CSS	-	Cascading Style Sheets
MySQL	-	My Structured Query Language
XHTML	-	eXtensible HyperText Markup Language
API	-	Application Programming Interface
ID	-	Identifikační číslo
SEO	-	Search Engine Optimization
URL	-	Uniform Resource Locator
DTD	-	Document Type Definition
AJAX	-	Asynchronous JavaScript and XML

Obsah

1. Úvod.....	6
1.1. Co je to SEO.....	7
2. Analýza projektu	8
2.1. Analýza webových nástrojů	8
2.2. Neimplementované atributy	9
2.3. Datová analýza	9
2.3.1. Specifikace.....	9
2.3.1. Lineární zápis	9
2.3.2. Datový slovník.....	10
2.3.3. Diagram práce robota	13
3. Testy a jejich parametry	14
3.1. Test dostupnosti stránky.....	14
3.2. MetaTagy	15
3.2.1. Titulek stránky	15
3.2.2. Klíčová slova	16
3.2.3. Popis stránky.....	16
3.3. Robots.txt	17
3.4. Validita stránky	17
3.5. Stránka na FaceBooku.....	18
3.6. Struktura nadpisů.....	19
3.7. Hustota klíčových slov	20
3.8. Google – pozice.....	21
3.9. Seznam - pozice	21
3.10. Google PageRank.....	22
3.11. S-Rank.....	22
3.12. Alexa Rank	23
3.13. Počet zpětných odkazů	23
3.14. Popularita stránek	23
3.15. Zaindexované stránky Google.com.....	24
3.16. Zaindexované stránky Seznam.cz	24
3.17. Google Trends.....	24
3.18. Seznam.cz: přesná shoda.....	25
3.19. Seznam.cz: rozšířená shoda.....	25
4. Druhy Testů	26
4.1. Běžný test.....	26
4.2. Testy robota	26
4.3. Testy trendů.....	26
5. Jak funguje Robot	27
6. Hodnotící algoritmus.....	28
6.1. Celkové hodnocení.....	28
6.2. Váhy a koeficienty	28
6.3. Sestavení vzorce pro výpočet.....	29
6.4. Nastavení vzorce	30
7. Problémy při realizaci.....	31
7.1. Ochrana proti robotům	31
7.2. Relevantnost výsledků.....	31
7.3. Google API.....	32

7.4. Spolehlivá práce robota	32
8. Statistické vyhodnocení.....	33
8.1. S-Rank	34
8.2. Alexa Rank	35
8.3. Google Rank vs. S-Rank	36
9. GUI	37
9.1. Grafický layout.....	37
9.2. Registrace a přihlášení	37
9.3. Projekty	38
9.4. Statistiky.....	38
9.5. Nástroje pro tvůrce stránek.....	39
9.5.1. Mapování a sitemapa stránky	40
9.5.2. Test validity a dostupnosti stránek	40
9.5.3. Zobrazení Lynx.....	40
9.6. Diskuze.....	41
9.7. Administrace	41
9.7.1. Bots.....	41
9.7.2. Trendy.....	42
9.7.3. Algoritmus.....	42
9.7.4. Filtry	42
9.7.5. Články.....	43
10. Závěr.....	44
11. Reference.....	45
PŘÍLOHY	46
I. Obsah přiloženého CD	46

Seznam tabulek

1	Srovnání současných českých SEO nástrojů.....	8
2	Lineární zápis.....	9
3	Zkrácený datový slovník projektu.....	10
4	Přehled testů a jejich hodnotících vah.....	28
5	Přehled vybraných koeficientů.....	30
6	Část výsledků statistických testů.....	33

Seznam obrázků

1	Schéma práce robota.....	13
2	Výpis výsledků dostupnosti a velikosti stránky na hlavní straně	14
3	Výpis výsledků Titulku stránky a Klíčových slov na hlavní straně.....	15
4	Výpis výsledků Popis stránky a Robots.txt na hlavní straně	17
5	Výstup W3 validátoru v okně webového prohlížeče (chyba).....	17
6	Výpis výsledků Validity a dostupnosti Stránky na FaceBooku na hlavní straně	18
7	Google.cz - Výsledky vyhledávání stránky Hoax.cz na stránkách facebook.com	19
8	Výpis výsledku testu Struktura nadpisů na hlavní straně	20
9	Výpis výsledku testu Klíčová slova na hlavní straně	20
10	Výpis výsledků pozice ve vyhledávači Google.com a Seznam.cz na hlavní straně	21
11	Výpis výsledků Google Rank, S-Rank, Alexa-Rank na hlavní straně.....	23
12	Výpis výsledků Počet zpětných odkazů a Popularity stránek na hlavní straně	23
13	Výpis výsledků indexované stránky, google trends, seznam shoda na hlavní straně	24
14	Formulář na hlavní straně projektu a zobrazení celkového výsledku	28
15	Omezení přístupu robotům Google.com.....	31
16	Seznam.cz: Graf S-Ranku v závislosti na pozicích ve výsledcích hledání	35
17	Hodnota Alexa Ranku je výrazně nižší u prvních 20 výsledků	35
18	Srovnání hodnot S-Rank a Google Rank	36
19	Vzhled webové aplikace	37
20	Registrace	37
21	Přihlášení	37
22	Výpis projektů	38
23	Ukázka možnosti nastavení filtru	38
24	Statistiky URL adresy zadané uživatelem	39
25	Výpis doplňujících nástrojů pro uživatele	40
26	Ukázka výpisu z prohlížeče Lynx.....	41
27	WYSIWYG editor v administraci pro psaní článků	43

Seznam výpisů zdrojového kódu

1	Ukázka zdrojového kódu definice meta tagu keywords.....	16
2	Správná definice nadpisů	19

1. Úvod

Internet se stal nedílnou součástí našich životů. Obsahuje miliony informací. Chceme-li nějaké informace najít, obvykle k tomu použijeme vyhledávač. Zadááme klíčová slova a vyhledávač nám zobrazí stránky, které hledané informace obsahují. Tvůrci webových stránek se z pochopitelných obchodních důvodů snaží, aby jejich stránky byly ve výsledcích nejlépe na první pozici. Soubor činností, o kterých se předpokládá, že mohou tuto pozici ovlivnit, se nazývá SEO¹ (Search Engine Optimization, česky Optimalizace pro vyhledávače).

Tato práce je zaměřena na úzkou část těchto činností. Pro měření úrovně SEO se používají obvykle webové nástroje, které podle určitých parametrů ohodnotí tuto úroveň. Tyto nástroje a jejich hodnocení SEO úrovně jsou založeny na určitých předpokladech, jejichž správnost není nikde dokázána.

Tato práce je zaměřena na vytvoření SEO nástroje, který SEO úroveň změří a vyhodnotí. Většina těchto nástrojů však povoluje otestovat v jednom kroku pouze jednu URL adresu. SEO analyzátor bude využívat robota, který bude automaticky procházet a testovat URL adresy registrovaného projektu a umožní uživateli zobrazit vývoj jednotlivých testovaných parametrů na časové ose. Robot bude využívat svých vlastností ke sběru dat, na kterých bude pomocí statistických metod ověřena závislost testovaných parametrů na pozici v nejvyužívanějších vyhledávačích Google.com² a Seznam.cz³. Tyto informace pak poslouží k nastavení hodnotícího algoritmu a budou zveřejňovány v pravidelných intervalech ve formě článku, který shrne zkoumané závislosti a upozorní na zajímavé souvislosti testovaných parametrů.

V kapitole 2, *Analýza webových nástrojů*, jsou srovnány nejpoužívanější české volně dostupné webové SEO nástroje. Na základě tohoto srovnání byly stanoveny testované parametry. Kapitola také obsahuje zjednodušenou analýzu projektu.

V následující kapitole 3, *Testy a jejich parametry*, budou jednotlivé vybrané parametry popsány, vysvětlen jejich význam a popsána logika algoritmu při získávání hodnot.

V kapitole 4, *Druhy testů*, budou popsány způsoby využití vlastností robota a logika získávání informací pro následné vyhodnocení.

V následující kapitole 5 je slovně popsán princip činnosti robota.

V kapitole 6, *Hodnotící algoritmus*, je popsán princip a samotný algoritmus, který vyhodnocuje dosaženou SEO úroveň.

V nejkratší kapitole 7 jsou stručně popsány problémy při realizaci a odůvodnění některých použitých technik.

Kapitola 8, *Statistické vyhodnocení*, popisuje výsledky zkoumání vlivu jednotlivých parametrů na pozici ve vyhledávačích.

Kapitola 9, *GUI*, popisuje použité grafické rozhraní a jednotlivé funkce dostupné uživateli a administrátorovi projektu.

¹ Někdy jsou tyto činnosti nazývány jako SEO optimalizace – což by ve výsledku znamenalo optimalizace pro vyhledávače (SEO) a její optimalizace, což je logicky špatně. Přesto je tento pojem takto hojně využíván.

² Základní adresou společnosti Google.com v této práci považujeme *www.google.com* i když ji lze libovolně zaměnit za *www.google.cz*, v některých částech textů pak obecně zmiňuji jen Google.

³ Dále v textu je zjednodušen název vyhledávače na *Seznam*

1.1. Co je to SEO

Pojem SEO (Search Engine Optimalization, optimalizace vyhledávacích strojů) označuje rozmanitou skupinu aktivit, které můžete vykonávat za účelem zvýšení cíleného provozu, jenž přichází na váš web z vyhledávacích strojů. Je možné, že jste se setkali i s dalšími názvy těchto aktivit – Search Engine Marketing (marketing vyhledávacích strojů), nebo Search Marketing (marketing vyhledávání). Tyto termíny zahrnují i věci, které děláte pro stránku samotnou (jako jsou změny v textu a HTML kódu). Dále zahrnují přímou komunikaci s vyhledávacími stroji a sledování dalších zdrojů provozu. Součástí SEO je také sledování, zkoumání a hodnocení konkurence.

SEO není reklama, i když může obsahovat reklamní komponenty. Nejedná se o PR (public relations, řízení vztahů s veřejností), i když může obsahovat úlohy komunikace podobné PR. Jedná se o oblast online marketingu, která se průběžně vyvíjí, takže může na první pohled působit komplikovaně. Ve svých základních cílech je ovšem velice jednoduchá. Cílem SEO je nárůst cílových návštěvníků. [5]

2. Analýza projektu

2.1. Analýza webových nástrojů

Tato kapitola popisuje základní charakteristiky volně dostupných webových nástrojů a jejich srovnání. Srovnávány byly české nástroje, které jsou uvedeny v Tabulce 1. Základní podmínkou pro zařazení nástroje do srovnání byla možnost bezplatného využití pomocí webového prohlížeče.

Parametr	Seo-pruvodce .cz	Nástroje-seo .eu	seo.wamos .cz	seo.oncz.net	seo-servis.cz
Čas zpracování	Ano	Ne	Ano	Ne	Ne
Meta Tagy	Ano	Ne	Ano	Ano	Ano
Titulek	Ano	Ne	Ano	Ano	Ano
Keywords	Ano	Ne	Ano	Ano	Ano
Description	Ano	Ne	Ano	Ano	Ano
DocType	Ano	Ne	Ano	Ano	Ne
Velikost stránky	Ano	Ne	Ano	Ne	Ano
Charset	Ano	Ne	Ano	Ano	Ano
Stáří domény	Ano	Ne	Ano	Ne	Ano
Robots.txt	Ano	Ne	Ano	Ano	Ano
Struktura H	Ano	Ne	Ano	Ano	Ano
DTD	Ano	Ne	Ne	Ano	Ano
Sémantika textu	Ano	Ne	Ano	Ne	Ano
Odstavce	Ano	Ne	Ne	Ne	Ano
Počet slov	Ano	Ano	Ne	Ne	Ano
Počet slov - dvoj a více fáze	Ne	Ano	Ano	Ne	Ano
% počtu klíčových slov	Ne	Ano	Ne	Ne	Ano
Zpětné odkazy Google	Ano	Ano	Ano	Ano	Ne
Zpětné odkazy Yahoo	Ano	Ano	Ano	Ano	Ano
S-Rank	Ano	Ano	Ano	Ano	Ano
Alexa Rank	Ano	Ano	Ano	Ano	Ano
Google Rank	Ano	Ano	Ano	Ano	Ano
Počet indexovaných stránek	Ano	Ano	Ano	Ano	Ano
Pozice Jyxo	Ano	Ano	Ano	Ne	Ne
Pozice Morfeo	Ano	Ano	Ano	Ne	Ne
Pozice Seznam	Ano	Ano	Ano	Ne	Ano
Pozice Google	Ano	Ano	Ano	Ne	Ano
Pozice Yahoo	Ano	Ne	Ano	Ne	Ne
Pozice Bing	Ano	Ne	Ano	Ne	Ne
Validity HTML a CSS	Ano	Ano	Ano	Ano	Ano
Hledanost výrazů	Ne	Ano	Ne	Ne	Ne
Popularita	Ne	Ne	Ano	Ne	Ano

Tabulka 1: Srovnání současných českých SEO nástrojů

Tabulka č. 1 shrnuje vlastnosti současných českých nástrojů a jejich srovnání. Červeně označené vlastnosti byly vybrány jako nejčastější nebo nejdůležitější a byly implementovány do SEO Analyzátoru.

2.2. Neimplementované atributy

I když některé zdroje v diskuzích uvádějí, že Doctype a DTD má svoji váhu, nakonec nebyly tyto parametry implementovány. Stáří domény obvykle v nástrojích nefunguje a nezobrazuje se korektně. A nepodařilo se najít efektivní způsob zjišťování, jelikož je tento údaj veden v evidenci *www.nic.cz* a informace z tohoto nástroje nejsou strojově dosažitelné.

Sémantika textu a rozeznání odstavců nebylo implementováno z důvodu optimalizace běhu robota, jehož čas by byl tímto neúměrně prodloužen.

Zpětné odkazy z Google.com nebyly implementovány z důvodu ochrany proti robotům a zjišťování hodnoty tohoto parametru by obnášelo další dotaz na servery Googlu a tím by docházelo k prodloužení času zpracování jednotlivých testů.

SEO Robot se zaměřuje na nejpoužívanější vyhledávače Seznam.cz a Google.com a z tohoto důvodu nebylo implementováno testování parametrů ve vyhledávačích Jyxo, Morfeo, Yahoo (kromě zpětných odkazů) a Bing.

2.3. Datová analýza

Ačkoliv je tato práce zaměřena na popis funkcí systému a vyhodnocení výsledků, je nutné si v několika málo krocích provést analýzu celé aplikace.

2.3.1. Specifikace

Mým úkolem je tvorba robota, který bude získávat výsledky testovaných parametrů a ve vhodné formě je bude prezentovat uživatelům. Vlastnosti robota pak budou využity k získání výsledků k analýze závislosti jednotlivých parametrů na pozici ve vyhledávači.

2.3.1. Lineární zápis

Následující tabulka reprezentuje lineární zápis. Primární klíče jsou vyznačeny tučně, cizí klíče jsou vyznačeny kurzívou.

seobot_activity	(<i>idact</i> , <i>idbot</i> , start, stop, pages)
seobot_admins	(<i>iduser</i> , login, heslo, email, activ, datum)
seobot_alg	(<i>idalg</i> , <i>idtest</i> , typ, koef, value1, value2)
seobot_bots	(<i>idbot</i> , namebot)
seobot_clanky	(<i>idclanku</i> , shorttext, nadpis, text, datum)
seobot_comments	(<i>id_comment</i> , <i>idforum</i> , typ, <i>id</i> , <i>id_user</i> , text, datum)
seobot_comments_activ	(<i>id_user</i> , hlavni, reakce, <i>idforum</i>)
seobot_comments_like	(<i>idlike</i> , <i>id_comment</i> , <i>id_user</i> , typ)

seobot_common_results	(id_c_result , url, keyword, time, google_pos, seznam_pos, g_rank, s_rank, a_rank, backlinks, popularity, g_index, s_index, s_rozsirena, s_presna, validity, density, h_structure, pagesize, meta, status, robots, facebook, vysledek)
seobot_diskuze	(idforum , nazev, autor)
seobot_engines	(idengine , engine)
seobot_filters	(idfilter , namefilter, popis, params)
seobot_online	(ip, username, time, team)
seobot_projects	(id_proj , id_user , url, keywords, last_check, filter, favourites, valid, hash)
seobot_results	(id_result , id_sm , id_proj , time, google_pos, seznam_pos, g_rank, s_rank, a_rank, backlinks, popularity, g_index, s_index, s_rozsirena, s_presna, validity, density, h_structure, pagesize, meta, status, robots, facebook, vysledek)
seobot_sitemaps	(id_sm , id_proj , url, ignored, status, last_check)
seobot_tests	(idtest , nametest, vaha)
seobot_trend_keys	(idtrendkey , trendkey, smazan)
seobot_trend_measure	(idmeasure , id_testname, idengine, idtrendkey, datum, no_results)
seobot_trend_results	(id_result , idmeasure , idtrendsite , time, google_pos, seznam_pos, g_rank, s_rank, a_rank, backlinks, popularity, g_index, s_index, s_rozsirena, s_presna, validity, density, h_structure, pagesize, meta, status, robots, facebook, vysledek)
seobot_trend_sites	(idtrendsite , idmeasure , position, url, checked)
seobot_trend_testnames	(idtestname , testname)
seobot_uzivatele	(iduser , login, heslo, email, activ, datum, foto, it_pozice)

Tabulka 2: Lineární zápis

2.3.2. Datový slovník

Datový slovník nám popisuje jednotlivé tabulky databáze a jejich atributy. Zkrácená tabulka také obsahuje typy atributů, jejich velikosti atd.

Název	typ	délka	klíč	null	index
seobot_activity					
idact	numeric	8	A	N	A
idbot	numeric	3	A	N	N
start	numeric	10	N	N	N
stop	numeric	10	N	N	N
pages	numeric	10	N	N	N
seobot_admins					
iduser	numeric	6	A	N	A
login	character	200	N	N	N
heslo	character	32	N	N	N
email	character	255	N	N	N
activ	character	2	N	N	N
datum	date		N	N	N
seobot_alg					

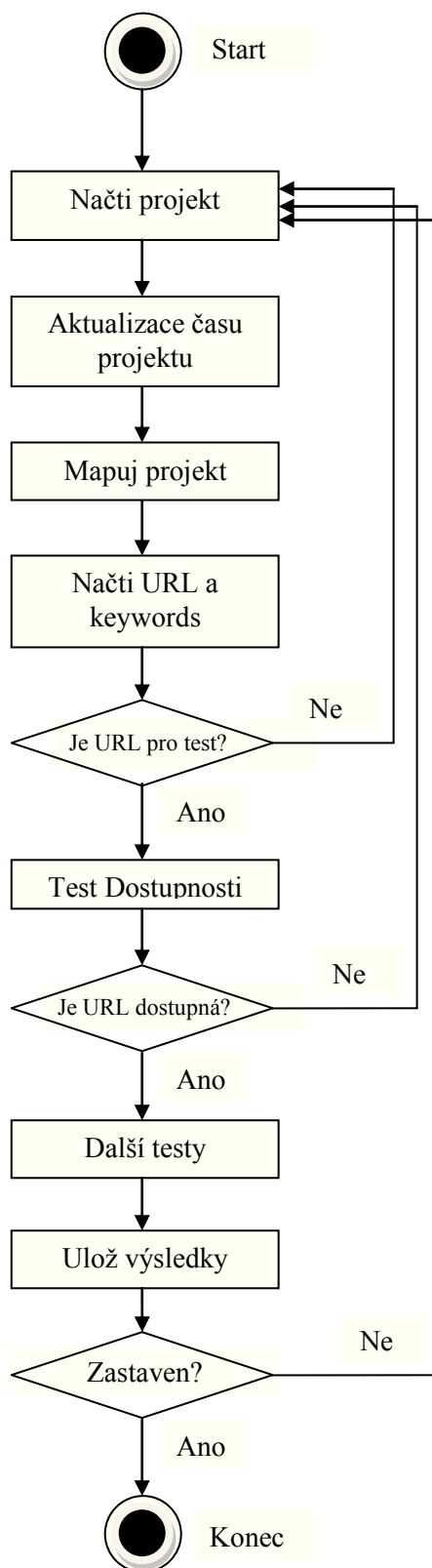
idalg	numeric	6	A	N	A
idtest	numeric	6	A	N	A
typ	character	10	N	N	N
koef	character	5	N	N	N
value1	character	10	N	N	N
value2	character	10	N	N	N
seobot_projects					
id_proj	numeric	6	A	N	A
id_user	numeric	6	A	N	N
url	text		N	N	N
keywords	text		N	N	N
last_check	numeric	10	N	N	N
filter	text		N	N	N
favourites	text		N	N	N
valid	character	2	N	N	N
hash	character	10	N	N	N
seobot_sitemaps					
id_sm	numeric	7	A	N	A
id_proj	numeric	6	A	N	N
url	text		N	N	N
ignored	character	1	N	N	N
status	character	50	N	N	N
last_check	numeric	10	N	N	N
seobot_tests					
idtest	numeric	6	A	N	A
nametest	character	200	N	N	N
vaha	character	5	N	N	N
seobot_trend_keys					
idtrendkey	numeric	3	A	N	A
trendkey	character	255	N	N	N
smazan	numeric	1	N	N	N
seobot_trend_measure					
idmeasure	numeric	5	A	N	A
id_testname	numeric	5	A	N	N
idengine	numeric	2	A	N	N
idtrendkey	numeric	3	A	N	N
datum	date		N	N	N
no_results	character	15	N	N	N
seobot_trend_results					
id_result	numeric	7	A	N	A
idmeasure	numeric	5	A	N	N
idtrendsite	numeric	6	A	N	N
time	numeric	10	N	N	N
google_pos	numeric	3	N	N	N
seznam_pos	numeric	3	N	N	N
g_rank	numeric	3	N	N	N
s_rank	numeric	3	N	N	N
a_rank	numeric	9	N	N	N
backlinks	numeric	9	N	N	N

popularity	numeric	6	N	N	N
g_index	numeric	6	N	N	N
s_index	numeric	6	N	N	N
s_rozsirena	numeric	6	N	N	N
s_presna	numeric	6	N	N	N
validty	character	5	N	N	N
density	character	20	N	N	N
h_structure	text		N	N	N
pagesize	character	6	N	N	N
meta	text		N	N	N
status	character	100	N	N	N
robots	text		N	N	N
facebook	character	100	N	N	N
vysledek	numeric	3	N	N	N
seobot_trend_sites					
idtrendsite	numeric	6	A	N	A
idmeasure	numeric	5	A	N	N
position	numeric	4	N	N	N
url	text		N	N	N
checked	numeric	10	N	N	N
seobot_trend_testnames					
idtestname	numeric	5	A	N	N
testname	character	150	N	N	N
seobot_uzivatele					
iduser	numeric	6	A	N	N
login	character	200	N	N	N
heslo	character	32	N	N	N
email	character	255	N	N	N
activ	character	2	N	N	N
datum	date		N	N	N
foto	character	10	N	N	N
it_pozice	character	50	N	N	N

Tabulka 3: Zkrácený datový slovník projektu

2.3.3. Diagram práce robota

Tento diagram práce robota zjednodušeně zobrazuje, jak probíhá samotná činnost robota. Některé činnosti byly v rámci zachování srozumitelnosti opomenuty.



Obrázek 1: Schéma práce robota

3. Testy a jejich parametry

3.1. Test dostupnosti stránky

Test dostupnosti stránky je spouštěn jako první a výsledky tohoto testu pak ovlivňují další chování celkového testu, nebo chování robota, který stránku testuje. Buď se zavolají funkce obsluhující další testy, nebo se vypíše hláška, že stránka není dostupná. V případě, že test volá robot tak se přejde k další testované stránce.

Tímto můžeme ušetřit spoustu času v případě, že stránka již neexistuje, nebo je v současné době nedostupná apod. Zároveň se pak vyhneme problémům při zpracovávání dalších testů, které nepředpokládají, že stránka neexistuje a navíc můžeme informovat uživatele, že stránka je již delší dobu nedostupná. Tímto monitoringem dostupnosti stránek⁴ tak uživatel získává nástroj, jehož vlastnosti lze nahradit použitím kombinací několika nástrojů od různých výrobců.

Vstupem tohoto testu je URL adresa testované stránky. Test stáhne obsah stránky a uloží ji do paměti. Po tomto stažení zjistí velikost zdrojového kódu. Pokud vše proběhne v pořádku tak výstupem tohoto testu je textový řetězec hodnot, oddělený "|". V případě testování z hlavní strany vrací hodnotu OK a velikost stránky. Do paměti (do SESSION⁵) pak ukládá znakovou sadu stránky, která se využívá pro další výpis a případné překódování výsledků dalších testů do jiné znakové sady. Zároveň uloží do SESSION načtený zdrojový kód a velikost stránky. V případě, že stránka není dostupná, výstupem tohoto testu je hodnota "-1".

Pokud je test spouštěn robotem pak výstup tohoto testu obsahuje další informace jako ID projektu, klíčová slova, která se načetla do SESSION společně s načtením URL adresy a dále jsou použita při dalších testech, nebo ID sitemapy⁶ (ID uložené stránky, která byla nalezena pomocí mapování⁷).

Testu	Výsledek testu
Dostupnost stránky:	Stránka je dostupná
Velikost stránky:	18 Kb

Obrázek 2: Výpis výsledků dostupnosti a velikosti stránky na hlavní straně

⁴ Monitoring dostupnosti stránek – soustavné sledování dostupnosti a funkčnosti URL adresy

⁵ SESSION – někdy také „sezení“ je proměnná, která může obsahovat libovolné hodnoty. Session se vytváří vždy pro konkrétního uživatele.

⁶ ID sitemapy – pořadové číslo URL adresy získané a uložené při mapování stránek, které tvoří sitemapu stránky

⁷ Mapování – funkce robota, který přistupuje na jednotlivé URL adresy, získává z nich další URL adresy, které nejsou uloženy v databázi

3.2. MetaTagy

Meta tagy jsou zvláštní značky, které se píšou do hlavičky stránky. V každé stránce je potřeba použít, meta tag pro kódování a description. Před několika lety byly metatagy velmi důležité. V rámci zdokonalování vyhledávacích algoritmů však jejich význam v současné době klesá.

Test *MetaTagy* získává informace o definici těchto metatagů a výstupem tohoto testu jsou textové řetězce oddělené "|", které jsou později podrobeny analýze.

3.2.1. Titulek stránky

Titulek je hlavní součástí výsledku ve vyhledávání. Ve zdrojovém kódu je reprezentován textem mezi HTML tagy⁸ <title> a </title>. Stručně několika slovy reprezentuje obsah stránky a tím dává uživateli najevo, zda stránka, kterou vyhledávač uživateli nabídl, opravdu popisuje problematiku nebo informace, které právě hledal. "Titulek stránky je velmi důležitý i z hlediska vzbuzení zájmu o vaši stránku ze strany hledajících." [1]

Algoritmus zpracovává text uložený v průběhu testu *Dostupnosti stránky*. Nezpracováváme však celý zdrojový kód, ale jen jeho část. Ze standardu HTML víme, že se titulek stránky musí nacházet v hlavičce stránky, resp. mezi párovým tagem <head> a proto vybereme část zdrojového kódu mezi těmito značkami a dále budeme pracovat pouze s tímto segmentem.

S touto vybranou pasáží budeme pracovat i v dalších testech. Na počátku algoritmu hledáme pozici tagu <title>. Pokud ji nenajdeme, předpokládáme, že titulek stránky není definován a výstupem je hodnota "-1". V opačném případě hledá ukončovací značku titulku </title> a zjistí její pozici. Následně je vybrán obsah mezi těmito značkami pomocí zjištěných pozic a přiřazen do proměnné, která je předána do výstupu testu.

Titulek stránky:	Tvorba www stránek, Eshopů, IS na míru, PC servis Neklikejte.cz
Klíčová slova stránky(meta):	Tvorba www stránek, Eshopy, Webové prezentace na míru, PC IT servis, Prodej IT techniky

Obrázek 3: Výpis výsledků Titulku stránky a Klíčových slov na hlavní straně

⁸ HTML Tagy – značky ve zdrojovém kódu, které dávají textu nějaký význam nebo formát, párové HTML tagy jsou reprezentovány dvojicí značek např. <a ...> a . První z nich znamená začátek tagu, druhá pak konec tagu. V HTML se používají i nepárové značky jako např.

3.2.2. Klíčová slova

O významu tohoto HTML tagu bylo napsáno příliš mnoho diskusních příspěvků. Jedna skupina uživatelů považuje tento HTML tag a jeho obsah za nutný a nepostradatelný, druhá skupina jej považuje za bezpředmětný a zbytečný. Pokud však můžeme touto značkou předat nějaké informace vyhledávači tak by měla být použita a její obsah musí být smysluplný.

Vstupem tohoto testu je, stejně jako v předchozím případě segment zdrojového kódu, který algoritmus přečte mezi HTML tagy <head> a </head>. Tato metaznačka se však může ve zdrojovém kódu objevovat v několika validních tvarech:

```
<meta name="keywords" content="... klíčová slova ..." /> nebo
```

```
<meta content="... klíčová slova ..." name="keywords" /> nebo
```

```
<meta http-equiv="Keywords" content="... klíčová slova ..."> nebo
```

```
<meta content="... klíčová slova ..." http-equiv="Keywords">
```

Výpis 1: Ukázka zdrojového kódu definice meta tagu keywords

Případně s ukončením ">" nebo pouhým ">" (bez lomítka)

Protože existuje několik variant jak zapsat tato klíčová slova, tak test nejprve najde slovo "keywords" uloží si pozici výskytu tohoto řetězce a od této pozice hledá ukončovací znak ">".

Pokud slovo "keywords" nenalezne, klíčová slova nejsou na stránce definována a na výstup vrací hodnotu "-1". Pokud ano, hledá od nalezené pozice slova "keywords" další slovo "content". Nejprve hledá napravo od nalezeného slova "keywords". Pokud nenalezne, hledá nalevo.

Content se nachází napravo od keywords:

V případě, nalezení slova "content" napravo od "keywords" test dále hledá rovnítko "=", za kterým se musí nacházet definovaná klíčová slova. Pak zjistí, zda se za rovnítkem nacházejí uvozovky, kterými jsou klíčová slova ohraničeny (tedy pokud jsou) a uloží si pozici.

Od této pozice pak hledá další uvozovky. Pokud nejsou uvozovky nalezeny tak hledá ukončovací znak ">". Tato pozice pak bude reprezentovat konec hledaného řetězce. Následně je vybrán text mezi uloženými pozicemi a vrací jej na výstup testu.

Content se nachází nalevo od keywords:

V tomto případě algoritmus hledá poslední slovo "content". Toto slovo je hledáno až do dříve uložené pozice dalšího slova "keywords". Od pozice, na které se nachází "content" pak algoritmus hledá pozici "name" nebo "http-equiv", což pro nás znamená konec hledaného řetězce

Následně test vybere text mezi uloženými pozicemi začátku a konce řetězce a obsah vrací na výstup testu pro metatag keywords. Popis tohoto algoritmu byl v rámci zachování srozumitelnosti zjednodušen.

3.2.3. Popis stránky

Obsah tohoto metatagu by měl několika větami popisovat obsah stránky. Měl by dávat smysl už jen proto, že vyhledávače jej používají při výpisu výsledků jako doplňující text k titulku stránky, neměl by obsahovat jen klíčová slova. Délka tohoto textu by měla odpovídat délce zobrazovaného textu ve

vyhledávačích (tedy 200-250 znaků). "V potaz ji vyhledávače berou zejména v okamžiku, kdy na samotné stránce není žádný jiný relevantní textový obsah, například když máte místo první stránky jen obrázek, rozcestník nebo animaci bez textů." [1] Získávání obsahu tohoto metatagu je stejné jako v bodě 2.2.

3.3. Robots.txt

Jednou z dalších možností jak mohou tvůrci webových stránek ovlivnit přístup robotů na jejich stránky je soubor robots.txt.

Tento soubor musí být umístěn v rootu webu⁹ a pomocí textu v tomto souboru lze např. zakázat určitému robotovi, aby navštěvoval některé části webu. Tuto možnost lze efektivně využít v případě, že chceme zabránit robotovi, aby indexoval například diskuze, které tvůrce stránek nemá pod plnou kontrolou. Mohou se v ní tedy objevovat informace, které nechceme, aby roboti viděli.

Vstupem do tohoto testu je testovaná URL adresa, ze které se vybere pouze základní tvar. Čímž dosáhneme root webu (např. <http://www.sport.cz/>). Samotný test, probíhá tak, že zkusí stáhnout obsah souboru robots.txt. Pokud existuje, vypíše i jeho obsah. Pokud se soubor nepodaří stáhnout, test vyhodnotí, že soubor robots.txt neexistuje.

Popis stránky(meta):	Neklikejte.cz se zabývá Tvorbou www stránek, Eshopů a informačních systémů na míru. Dále zprostředkováváme nákup IT techniky a poskytujeme PC servis
Robots.txt:	Existuje User-agent * Disallow /stats/ Sitemap: http://www.neklikejte.cz/sitemap.xml

Obrázek 4: Výpis výsledků Popis stránky a Robots.txt na hlavní straně

3.4. Validita stránky

Validita stránek znamená, že zdrojový kód stránek je psán dle platných a stanovených standardů. Tyto standardy byly definovány konsorciem W3. Toto konsorcium na svých stránkách provozuje validátor, pomocí kterého si můžou tvůrci ověřit, zda jsou stránky validní nebo nikoliv. Validátor lze vyzkoušet na adrese <http://validator.w3c.org>

Error found while checking this document as XHTML 1.0 Transitional!	
Result:	1 Error
Address:	http://www.neklikejte.cz/
Encoding:	windows-1250 (detect automatically)
Doctype:	XHTML 1.0 Transitional (detect automatically)
Root Element:	html
Root Namespace:	http://www.w3.org/1999/xhtml



The W3C validators rely on community support for hosting and development. [Donate](#) and help us build better tools for a better web.



Obrázek 5: Výstup W3 validátoru v okně webového prohlížeče (chyba)

⁹ Root webu – je kořenový adresář, ve kterém je umístěn hlavní soubor index.html, (případně .php, .asp a jiné)

O významu validity stránek pro vyhledávače se do jisté míry vedou diskuze dodnes. Můj závěr je takový, že validita stránek velký význam pro vyhledávače nemá. Obzvlášť když vyhledávací roboti dokážou chyby ignorovat. Na druhou stranu to vypovídá o tvůrci webových stránek a jeho "čisté" práci.

Pokud se na stránce nevyskytnou kritické chyby, tak jejich přítomnost na stránce není ničím škodlivým. Ovšem pokud se na stránce vyskytnou chyby, může se stát, že robot může přečíst stránku špatně a můžeme se tak připravit o možnost předat vyhledávači důležité informace. Proto jsem tento test do výzkumu nakonec zařadil.

Samotný test využívá výše zmíněné služby validátoru. Vstupem tohoto testu je opět URL testované stránky, která je předána do validátoru, který vyhodnotí, zda jsou stránky validní a pokud ne tak kolik chyb obsahují. Tato informace je pak testem přečtena a předána na jeho výstup jako číselná hodnota. Pokud se test provádí z hlavní strany projektu pak je v případě nevalidity stránky zobrazen odkaz na validátor, po jeho rozkliknutí se zobrazí chyby, které se na stránce vyskytnou. Odkaz není dostupný na jiných místech projektu, aby nedošlo k mylnému informování uživatele, v případě, že by prohlížel výsledky s časovým odstupem. Mohla by nastat situace, kdy systém vykazuje chyby, které již aktuálně neexistují.

W3 Validity:	1 chyb --> Odkaz
Stránka na FaceBooku:	Nalezeno --> Odkaz

Obrázek 6: Výpis výsledků Validity a dostupnosti Stránky na FaceBooku na hlavní straně

3.5. Stránka na FaceBooku

Vzhledem k neustále rostoucímu vlivu FaceBooku ve světě internetu a médií, začíná být vlastnictví stránky na FaceBooku nutností. Přestože nelze jednoznačně zjistit, zda Facebook stránka webového projektu na FaceBooku opravdu existuje, je tomuto testu ve výsledku přiřazována minimální váha.

Pokud si na FaceBooku majitel stránky založí stránku o svém webovém projektu a v názvu použije URL adresu projektu, může odkaz dostupný ve výsledcích vyhledávače vypadat např. takto: www.facebook.com/pages/HOAXCZ/10150112729530615 - URL adresa na FaceBooku neobsahuje tečky.

Protože jsou stránky takovýmto způsobem dohledatelné, využije test dotazu na Google. Vstupem testu je URL adresa v upraveném základním tvaru - tedy jen název domény bez tečky a koncovka. Výsledkem je pak hodnota 0, pokud stránka ve výsledcích nebyla nalezena, hodnota 1 v opačném případě.

Tyto výsledky však nemusí být relevantní. Prodejce sportovní obuvi značky XY může mít pro své produkty založenu stránku na FaceBooku s názvem např. "Sportovní boty pro vrcholové sportovce". Tato stránka může být prezentační stránkou pro produkty XY, přitom název stránky neobsahuje název URL adresy.

Proto byla tomuto testu přiřazena minimální váha a má sloužit jako vodítko pro majitele webových projektů v případě, že stránku na FaceBooku ještě nemají, aby nesporných reklamních výhod FaceBooku využili anebo nad touto možností alespoň přemýšleli.



Obrázek 7: Google.cz - Výsledky vyhledávání stránky Hoax.cz na stránkách facebook.com

3.6. Struktura nadpisů

Tak jako titulek stránky stručně popisuje obsah stránky, tak nadpisy popisují obsah textu pod nimi. Měly by stručně charakterizovat text v následujícím odstavci. Vyhledávače pak těmto nadpisům přikládají velkou váhu, a proto tento test kontroluje, zda jsou nadpisy použity a zda jsou použity správně. Správně použité nadpisy mají např. následující strukturu:

<H1>Diplomová práce SEO analyzátor</H1>

<H2>Jednotlivé testy</H2>

<H3>Test dostupnosti stránky</H3>

Výpis 2: Správná definice nadpisů

Tyto nadpisy v podstatě reprezentují rozdělení knihy do kapitol a podkapitol. Nadpisů lze použít více (H4, H5), ale tyto ve většině případů nejsou použity a proto nejsou testovány.

Vstupem testu je zdrojový kód stránky, kde v první části testu je vyhledán hlavní nadpis (H1) a uložen text uvnitř značek <H1> a </H1>, zároveň se u tohoto textu zjišťuje přítomnost klíčového slova. Pokud se klíčové slovo v textu hlavního nadpisu vyskytuje, je uložena jeho pozice. Toto slouží k analýze, zda mají vyšší účinnost nadpisy, které obsahují klíčové slovo na začátku textu nadpisu nebo zda na pozici klíčového slova nezávisí.

Následně test ověří, zda se od pozice konce prvního hlavního nadpisu <H1> nenachází další hlavní nadpis <H1>. Což by bylo špatně a nebyla by splněna validita stránky (viz. kapitola 3.4. *Validita stránky*). Pokud ano tak se na výstup testu pošle hláška, že stránka obsahuje více nadpisů H1.

V další části testu jsou hledány nadpisy H2, kterých může být neomezené množství. Po vyhledání je uložen jejich text a zjištěna pozice klíčového slova. Protože za nadpisy H2 mohou následovat nadpisy H3, je tato skutečnost ověřena, pokud ano, jsou uloženy texty a pozice klíčového slova i nadpisů H3. Tento krok se opakuje až do té doby, dokud test nenajde již další nadpis H2 nebo H3 nebo pokud narazí na konec stránky (zdrojového kódu).

Do funkce, která ukládá výsledky do databáze je posílán řetězec hodnot. První hodnota vyjadřuje, zda je struktura nadpisů v pořádku. Pokud ano, vrací hodnotu OK, pokud není v pořádku, vrací text, který popisuje chybu - např. "Není H2". Ve výstupním řetězci testu pak následuje textový

výpis jednotlivých nadpisů a pozic klíčových slov, pokud se v nadpisech nacházejí. V případě, že se klíčová slova nenacházejí v textu nadpisu, tak test vrátí místo pozice hodnotu "-1".

Struktura nadpisů (pozice klíčového slova):	<p>Struktura nadpisů je správná</p> <p>Text H1: (1x)</p> <ul style="list-style-type: none"> - Tvorba www na míru NEKLÍKEJTE.CZ (-1) <p>Text H2: (3x)</p> <ul style="list-style-type: none"> - Webové stránky, Eshopy na míru, PC servis (-1) - ZPRÁVY: Vždy stojíme na Vaší straně (-1) - Naše nabídka (-1)
---	---

Obrázek 8: Výpis výsledku testu Struktura nadpisů na hlavní straně

3.7. Hustota klíčových slov

Hustota klíčových slov vyjadřuje poměr klíčového výrazu k celkovému počtu slov na stránce. V obecných diskuzích se toto téma opět stává poněkud kontroverzním. Ne však ve smyslu významu hustoty klíčových slov, ale ve smyslu výše hustoty klíčových slov.

Obecně se hustota vypočítává jako procentuální vyjádření počtu slov reprezentující klíčové slovo k celkovému počtu slov obsaženém na stránce. Tedy obsahuje-li stránka celkem 100 slov a klíčové slovo se zde vyskytuje 3x pak je hustota klíčových slov 3%.

Mnozí označují optimální hodnotu mezi 2-5%. V diskuzích a článcích pak nalezneme názory, které tvrdí, že optimální hodnota je až 10%. Problémem stanovení optimální hodnoty je pak skutečnost, že v případě nízké hodnoty může být vliv hustoty klíčových slov podhodnocen. V opačném případě, kdy je hustota klíčových slov příliš vysoká, může být bráno použití klíčového slova jako SPAM. Hodnocení hustoty jako SPAM je logické. Nezřídka kdy jsem narazil na stránky, které obsahovaly mnoho klíčových slov za sebou a v nelogickém použití. Takto "přeoptimalizovaný" text byl téměř nepoužitelný a špatně čitelný.

Ačkoliv je samotný výpočet triviální, získání čistého textu z html kódu už tak jednoduché není. Zejména proto, že kód obsahuje spoustu textu v popiscích obrázků, popisovačů odkazů, komentářích apod. Pro tento účel bylo použito knihovny ze stránek: <http://scripts.ringsworld.com/development-tools/class.html2text/class.html2text.inc.html>

Samotný test je pak konstruován tak, že mu předáme URL adresu, funkce v knihovně pak převede zdrojový kód na čistý text a rozdělí slova do pole dle počtu. Testu je pak předán další vstupní parametr - klíčové slovo. Test ověří, zda se zadané klíčové slovo nachází v poli, které vytvořila knihovna.

Požadovaným výsledkem testu je celkový počet slov na stránce a hodnota hustoty klíčového slova. Problémem této použité knihovny je, že nedokáže rozpoznat různé tvary slov: např. bíl-ý, bíl-á, bíl-é apod. a nedokáže analyzovat dvouslovná (a víceslovná) klíčová slova. Pro hlubší testování vlivu hustoty klíčových slov na pozici ve vyhledávacích tak budou použita jednoslovná klíčová slova v nejběžnější se vyskytující tvarech, aby bylo dosaženo co nejpřesnějších výsledků.

Keyword Density:	Počet slov: 268 Klíčové slovo nenalezeno
------------------	---

Obrázek 9: Výpis výsledku testu Klíčová slova na hlavní straně

3.8. Google – pozice

Google.com je celosvětově nejúspěšnější vyhledávač na internetu. Na českém trhu však neustále svádí boj o pozici jedničky s tuzemským vyhledávačem Seznam.cz - podle některých pohledů se již jedničkou stal. Od českého vyhledávače se vyznačuje vyšším počtem výsledků, menší mírou komerčních odkazů a mnozí jej využívají pro jeho vyhledávání, které považují za relevantnější. Nespornou výhodou tohoto vyhledávače je množství aplikací a nástrojů, které poskytuje svým uživatelům zpravidla zdarma.

Samotný test zjišťuje, zda se URL adresa nalézá ve výsledcích vyhledávání na vstupní klíčový výraz. Pokud je URL ve výsledcích nalezena, vypíše pozici, na které se nalézá. Test přistupuje na stránku Googlu, obsahující 100 výsledků vyhledávání na klíčový výraz, na které pak hledá vstupní URL adresu. Pokud nenalezne, přechází na další stránku. Takto je výsledek hledán až do určitého počtu stránek, který lze nastavit. Ze zkušenosti a vzhledem k optimalizaci provozu robota bylo nastaveno, že tento test bude vyhledávat maximálně do 5. strany tedy do 500. výsledku.

3.9. Seznam - pozice

Jak již bylo zmíněno, vyhledávač Seznam.cz je jedním ze dvou nejpoužívanějších vyhledávačů na českém trhu. Další vyhledávače jsou využívány v menší míře. Proto se stává cílem tvůrců stránek a odborníků na optimalizaci dosáhnout co nejlepšího umístění ve výsledcích vyhledávání právě těchto dvou nejpoužívanějších vyhledávačů.

Test podobně jako v předchozím případě přistupuje na stránky Seznam.cz s výsledky vyhledávání na vstupní klíčový výraz a hledá vstupní URL adresu. Vyhledávač Seznam.cz neumožňuje vypsát více než 20 výsledků na stránku. Test taktéž prohledává až do 5té stránky, proto se ověřuje prvních 100 výsledků.

Výstupem testu je buď hodnota "-1" pokud nebyl výsledek nalezen, v případě, že nalezen byl, tak vrací pozici vstupní URL adresy. V případě, že je test volán z hlavní strany projektu tak je ve výsledku zobrazen i odkaz na výsledky vyhledávání a uživatel tak může ověřit, které stránky se vyskytují v okolí umístění jeho stránky. Jak již bylo zmíněno u validace v kapitole 3.4 *Validita stránky*, v dalších částech systému již tento odkaz není aktivní.

Google:	222 --> Odkaz
Seznam:	Nenalezeno

Obrázek 10: Výpis výsledků pozice ve vyhledávači Google.com a Seznam.cz na hlavní straně

3.10. Google PageRank

PageRank (dále jen PR) jak již název napovídá je údaj, který číselně vyjadřuje hodnocení stránky. Jedná se o hodnocení důležitosti stránky. Vychází z jednoduchého principu počtu odkazů a jejich důležitosti. Tzn., pokud na stránku X odkazují stránky s vysokou důležitostí, předpokládá se, že stránka X má také vysokou důležitost.

Interval PR není veřejně znám, jedná se ale o skalární veličinu. Každá URL má jinou hodnotu PR. Je možné, že PR dosahuje hodnot např. od miliontin do miliard. Pro zjednodušení a možná i kvůli utajení se přesné hodnoty nezveřejňují. V současné době se PR zaměňuje za GTPR (Google Toolbar PageRank). [1]

Jelikož přesnou hodnotu nelze zjistit, pracuje i SEO analyzátor s hodnotou GTPR. Výstupem tohoto testu je tedy přepočítaný PR, který je zveřejňován a nabývá hodnot od 0 do 10. Přičemž dosažení hodnoty 10 je velice těžké a na internetu dosahuje tohoto hodnocení velice málo stránek. Existuje více vzorců na výpočet PR. Jedním z možných je:

$$PR(A) = (1-d)/m + d * (PR(T1)/C(T1)) + \dots + (PR(Tn)/C(Tn))$$

Vysvětlení:

- PR(A) – PageRank stránky A
- T1 až Tn – stránky, které odkazují na stránku A
- C(T) – počet odkazů vedoucích ze stránky T
- d – tlumící (dampening) faktor
- m – celkový počet oindexovaných stránek
- Vstupní hodnoty PR(Ti) jsou hodnoty z posledního přepočítávání PR

Znění vzorce čerpáno z [1]

Skriptů, které zjišťují hodnotu PR určité URL stránky, existuje několik. Pro potřeby SEO analyzátoru byl použit skript, autora Emre Odabase, dostupný na stránkách <http://zfexts.googlecode.com/svn-history/r2/trunk/Gdata/GooglePR.php>

3.11. S-Rank

Stejně jako Google tak i Seznam používá svůj PageRank. Dříve dosahoval S-Rank hodnot od 0 do 100. Nedávno však tvůrci Seznamu přešli na ohodnocení, které dosahuje hodnot 0 až 10. SEO analyzátor však násobí získanou hodnotu 10x kvůli jednoznačnému odlišení hodnot Google PageRanku a S-Ranku. Tato upravená hodnota je použita také proto, že většina uživatelů, kteří tuto hodnotu sledují, jsou zvyklí na hodnocení v rozsahu 0-100.

Skript získávající hodnotu S-Rank je použit ze stránky <http://xrank.cz/download#licence>, která tuto verzi poskytuje volně k použití, jedinou podmínkou používání je umístění odkazu na web projektu XRank.cz.

3.12. Alexa Rank

Alexa rank vyvinula kalifornská společnost Alexa Internet. Na rozdíl od předešlých ranků jeho hodnotu neovlivňuje počet vedoucích odkazů na stránku, ale počet návštěvníků. Nezáleží ale jen na samotném počtu návštěvníků. Návštěvníci stránek musí mít ještě nainstalovaný Alexa toolbar. Alexa rank má hodnoty od 1 do několika milionů (cca 25mil). Čím nižší Alexa rank web má, tím více návštěvníků s nainstalovaným Alexa toolbarem na něj zavítalo. Text přebrán z [1].

Vzhledem k tomu, že většina obyčejných návštěvníků v České republice a tedy i potenciálních klientů provozovatelů webových stránek Alexa Rank neznají a nemají nainstalovaný Alexa toolbar, bude tomuto testu snížena váha a tím se méně projeví tato hodnota v celkovém hodnocení stránky.

Google Pagerank	0
S-Rank	60 / 100
Alexa-Rank	0

Obrázek 11: Výpis výsledků Google Rank, S-Rank, Alexa-Rank na hlavní straně

3.13. Počet zpětných odkazů

Tento test zjišťuje počet odkazů, které vedou na testovanou URL adresu. Počet stránek, jak již bylo popsáno výše, příznivě ovlivňuje PageRanky. Mimo to má další význam. Čím více stránek odkazuje na určitou webovou stránku, tím více mají možností potenciální návštěvníci najít web. Návštěvníci tak získávají více možností jak zjistit, že web existuje, což je pro tvůrce webových stránek velice důležité.

Vstupem tohoto testu je URL adresa. Výstupem pak počet stránek, které odkazují na tuto URL adresu. Test pošle dotaz na stránku z <http://search.yahooapis.com/>, která obsahuje seznam odkazujících stránek na URL adresu a také celkový počet těchto stránek. Odpověď dotazu je ve formě zdrojového kódu stránky, ze kterého je přečtena informace o počtu stránek a předána na výstup testu.

3.14. Popularita stránek

Popularita stránek je podobná počtu zpětných odkazů. Přesto se odlišuje v jedné zásadní věci. Zatímco počet zpětných odkazů je počet odkazů, na který lze skutečně kliknout a dostat se tak na určenou URL adresu, popularita stránek je počet všech výskytů URL adresy, tedy je započítána i jakákoliv zmínka v prostém textu.

Vstupem tohoto testu je opět URL adresa. Výstupem pak počet stránek, na kterých je vstupní URL adresa zmíněna. Pro zjištění tohoto testu je použit výstup z vyhledávání Googlu.

Počet zpětných odkazů:	570
Popularita stránek:	2770 Počet výskytů mezi webovými stránkami

Obrázek 12: Výpis výsledků Počet zpětných odkazů a Popularity stránek na hlavní straně

3.15. Zaindexované stránky Google.com

Indexace je činnost vyhledávače (nebo jeho části - robota) při které vyhledávač prochází jednotlivé stránky a stahuje jejich textový obsah, který si uloží do databáze a je dále zpracováván. Snahou tvůrců webových stránek je, aby vyhledávač měl informace o všech podstránkách webu, který vytvořili. Tím se zajistí, že vyhledávač má povědomí o tom, co se přesně na stránkách nachází a tím je větší možnost, že zobrazí stránky ve výsledcích vyhledávání. Jednoduše řečeno, o čem vyhledávač neví (nemá zaindexováno), nemůže nabídnout ve výsledcích.

Vstupem tohoto testu je opět URL adresa, která je předána vyhledávání na Googlu takovým způsobem, že Google zobrazí stránky na dané doméně. Výstupem tohoto testu je pak celkový počet takto nalezených stránek.

3.16. Zaindexované stránky Seznam.cz

Stejně tak jako Google i Seznam indexuje stránky. Test funguje naprosto stejně jako u Googlu, s tím rozdílem, že dotaz je posílán na Seznam.cz.

Zaindexované stránky Google:	51
Zaindexované stránky Seznam:	43
Google trends:	Google trends na slovo tvorba www
Seznam rozšířená shoda (hledanost):	Za den: 1 Za týden: 1
Seznam přesná shoda (hledanost):	Za den: -- Za týden: --

Obrázek 13: Výpis výsledků indexované stránky, google trends, seznam shoda na hlavní straně

3.17. Google Trends

Následující 3 nástroje neplní funkci testů, jsou to spíše nástroje pro uživatele. Google Trends je nástroj od společnosti Google, který na zadané klíčové slovo zobrazuje vývoj hledanosti klíčového výrazu v čase. Nástroj poskytne uživateli odkaz na tento přehled přímo na stránkách Googlu.

Google trends zobrazuje graf hledanosti klíčového slova v jednotlivých oblastech (regionech), městech a v určitých jazycích. I když tento nástroj neukazuje přímo hodnoty hledanosti, lze tak z grafu vyčíst například sezónnost hledání. Např. u klíčového slova lyže, lze z grafu vyčíst, že toto slovo uživatelé hledají se vzrůstající tendencí a vrcholem těsně před Vánoci. Pak hledanost tohoto slova klesá.

Při správném využití těchto informací pak lze zacílit např. prodej výrobků, připravit se na vyšší počet objednávek nebo vhodně naplánovat akční slevy. Tento nástroj se pak stává velice užitečným pro marketingové oddělení firmy.

3.18. Seznam.cz: přesná shoda

Seznam také poskytuje informace o hledanosti klíčového slova. Zobrazuje číselné informace za poslední období (obvykle za poslední dva měsíce). Tedy čísla, která vyjadřují počet hledání klíčového slova za den a týden.

Český vyhledávač poskytuje celkem tři hodnoty. Maximální četnost hledání, minimální četnost a průměrná hodnota. Právě tato průměrná hodnota hledanosti vstupního klíčového výrazu je uživateli zobrazena. Zobrazeny jsou pouze hodnoty, u kterých je dostatečný počet výsledků – zpravidla se jedná o číselné hodnoty, které jsou větší než 100. Toto omezení je závislé na vyhledávači.

Přesnou shodu Seznam v nápovědě definuje takto [2]:

- porovnávání s dotazem se klíčové slovo neskloňuje, nečasuje a neodstraňuje se mu diakritika.
- dotaz musí být v přesném znění klíčového slova, aby došlo k zobrazení výsledku

3.19. Seznam.cz: rozšířená shoda

Jedná se o stejný nástroj jako v předchozí kapitole, ovšem tentokrát jsou zobrazeny údaje o hledanosti slova s rozšířenou shodou. Seznam definuje v nápovědě rozšířenou neboli volnou shodu takto [2]:

- při porovnávání s dotazem se klíčové slovo skloňuje, časuje a odstraňuje se mu diakritika.
- v případě dotazu nezáleží na pořadí slov víceslovného klíčového slova, ale na jejich výskytu v dotazu.

4. Druhy Testů

V předchozí kapitole byly popsány jednotlivé parametry testů. Testy byly implementovány tak, aby bylo možné využít jejich vlastností v několika druzích celkových testů.

4.1. Běžný test

Běžný test je prováděn uživatelem z *Hlavní stránky* projektu. Zde je možné otestovat pouze jednotlivé stránky. Výsledky těchto testů jsou ukládány, protože obsahují další možnost získání cenných informací k pozdější analýze. Uživatel tímto testem dostane k dispozici kompletní výsledky a zároveň se v pravé části zobrazí odkaz na jeho webovou stránku. Získá tím krátkodobě zpětný odkaz ze stránek projektu.

4.2. Testy robota

Jednou z hlavních myšlenek tohoto projektu je, že registrovaným uživatelům budou prováděny testy automaticky a to na všech podstránkách jejich projektu. Robot za svého běhu neustále testuje jednotlivé projekty a jejich jednotlivé URL adresy získané při automatickém mapování. Výsledky po každém provedeném testu uloží a lze tak získat historický přehled vývoje jednotlivých parametrů.

4.3. Testy trendů

Hlavním zdrojem informací pro hodnocení vlivu jednotlivých testovaných parametrů na pozici stránky ve vyhledávačích jsou však testy trendů a výsledky z nich získané. Tyto testy využívají opačné myšlenky předchozích testů. Zatímco výše popsané testy získávají hodnoty parametrů jednotlivých stránek a snaží se je vyhodnotit, testy trendů se zaměřují na prvních 100 výsledků z každého vyhledávače a zkoumají jejich vlastnosti a souvislosti.

Tyto testy jsou připravovány v administraci¹⁰. Pro daná klíčová slova se pomocí připraveného algoritmu uloží URL adresy prvních 100 výsledků z vyhledávače Google a Seznam a na těchto URL adresách se provádějí jednotlivé testy. Takto získané výsledky jsou pak podrobeny důkladné analýze.

Výhodou těchto testů je možnost zjistit okamžitý výsledek. Lze tak i vypořizovat jednotlivé trendy vyhledávačů. Tedy z analýzy plyne, zda a na které parametry kladou vyhledávače větší důraz, nebo v jakých hladinách se vyskytují parametry stránek umístěných na prvních pozicích.

¹⁰ Administrace – je myšlena administrace projektu. Této části projektu je věnována jedna z dalších kapitol.

5. Jak funguje Robot

U každého projektu je evidováno ID projektu, URL adresa, klíčová slova, na které je testována a čas posledního testu. Při startu robota a při každém dalším průchodu se dle času načte projekt, který byl nejdelší dobu netestován a tento čas se aktualizuje. Robot si uloží do paměti ID projektu a to pak předává dalším částem algoritmu.

Následně se spustí mapování, při kterém se načte stránka vybraného projektu, která ještě nebyla zkontrolována nebo ta která byla nejdelší dobu netestována. Při mapování se na stránce vyhledávají odkazy a ukládají do databáze. Pokud odkaz vede na testovanou doménu, nastaví se u ní parametr Ignored na 0. Pokud stránka odkazuje na "cizí" doménu je odkaz označen jako Ignored nebo-li parametr Ignored je nastaven na hodnotu 1. Pokud již v databázi odkaz u projektu existuje tak se neukládá. Stránka se zařadí mezi ignorované také v případě, že odkaz míří na soubor s nepovolenou příponou - mezi povolené přípony patří: .php, .php3, .html, .htm, .xhtml, .asp, .aspx. Ignorovány jsou také stránky, které nejsou dostupné na těchto doménách 1.stupně: .cz, .sk, .eu.

Tento způsob nám zaručuje několik výhodných vlastností. Jednou z nich je spravedlivé testování jednotlivých projektů a jejich pravidelné střídání. Zároveň tak dochází k automatickému mapování jednotlivých stránek a postupně se vytvářejí jednotlivé sitemapy. Další nespornou výhodou tohoto způsobu načítání projektů je přístup k databázi projektů a stránek, k níž může přistupovat více instancí robota. Každý robot si tak načte svůj projekt, kterému se bude věnovat. Tímto se dá efektivně zkrátit čas zpracování kontrol jednotlivých stránek.

Po ukončení mapování je načtena stránka vybraného projektu, která ještě nebyla testována nebo stránka, která byla testována minimálně před 24h. Pokud se taková stránka v databázi nenachází, skript vrací příkaz, kterým se robot vrátí na začátek a znovu načítá další projekt.

Tímto se vyhneme několika výsledkům během jediného dne, které bývají obvykle duplicitní a byly by zbytečné. Přesto se však automaticky provede mapování a tím se naleznou případné nové stránky, které se na stránce mohly od poslední návštěvy robota objevit.

Dalším krokem robota je samotné testování. Prvním testem, který je spuštěn, je *Test dostupnosti a velikosti stránky*. Odkaz, který byl při mapování nalezen, totiž nemusí již na stránce existovat, nebo může vést na neexistující stránku. Pokud je test vyhodnocen jako úspěšný, výsledky se uloží a spouští se další testy. Pokud test vyhodnotí stránku jako nedostupnou, uloží chybu a u stránky se uloží záznam, že stránka byla nedostupná. Pokud se takto stránka vyhodnotí po třetí, je následně označena jako ignorovaná. Jak již bylo zmíněno v kapitole 3.1. *Dostupnost stránky*, je-li stránka nedostupná, další testy již nejsou spouštěny a robot se opět vrací na začátek a načítá další projekt.

Protože vyhledávače Google a Seznam disponují ochranou proti robotům je nutné testy načítající data z těchto vyhledávačů spouštět za sebou a s časovým odstupem několika sekund. Ostatní testy jsou spouštěny najednou. Výsledky z jednotlivých testů jsou průběžně ukládány do paměti. Jakmile jsou k dispozici všechny výsledky, robot je uloží a označí URL stránku časem poslední kontroly. Pak opět robot pokračuje od začátku, načte další projekt a následuje výše popsany postup.

6. Hodnotící algoritmus

Jednotlivé hodnotící testy byly implementovány tak, aby jejich výstupem byly číselné hodnoty. V případě textových výsledků (titulek stránky, metatagy apod.) převedeme text na vhodnou číselnou hodnotu, např. délka řetězce, počet řetězců, pozice klíčového slova, vyplněno/nevyplněno (0,1) apod. Takto máme k dispozici sadu číselných výsledků, které můžeme vyhodnotit.

6.1. Celkové hodnocení

Pro jednodušší vyhodnocování jsem si určil maximum, které může webová stránka dosáhnout. Toto maximum je nastaveno na 10 000 bodů. Tento výsledek se podělí 100 a zaokrouhlí na celá čísla nahoru a tím můžeme web ohodnotit v rozmezí od 0-100 bodů.

Uživateli, který testuje své webové stránky z hlavní strany projektu je výsledek zobrazen ihned po dokončení testování. Výsledky jsou při ukládání ohodnoceny hodnotícím algoritmem, na jehož výstupu je vrácena hodnota, jak bylo popsáno v předchozím odstavci, od 0 do 100 a ta je následně uživateli zobrazena.

Zjistěte si pozice a ranky svých stránek (v 1.138)

Doposud provedeno 296 měření

Zadejte URL bez zpětného lomítka

URL adresa:

Klíčové slovo:

Start

30

Obrázek 14: Formulář na hlavní straně projektu a zobrazení celkového výsledku

6.2. Váhy a koeficienty

Protože každý testovaný parametr ovlivňuje stránku jiným způsobem, byly jednotlivým testům přiřazeny tzv. váhy. Např. titulek stránky je pro vyhledávače důležitý, proto má výsledek tohoto testu vysokou váhu. Naproti tomu, zda má projekt stránku na FaceBooku nelze s jistotou zjistit, proto výsledku tohoto testu byla přiřazena váha minimální.

Test	Váha
Velikost stránky	0.5
Titulek stránky	5
Klíčová slova stránky(meta)	2
Popis stránky(meta)	2
Robots.txt	1.5
W3 Validity	1.5
Stránka na FaceBooku	0.4

Struktura nadpisů (pozice klíčového slova)	4
Keyword Density	4
Google	5
Seznam	5
Google Pagerank	5
S-Rank	5
Alexa-Rank	1
Počet zpětných odkazů	5
Popularita stránek	4
Zaindexované stránky Google	3
Zaindexované stránky Seznam	3

Tabulka 4: Přehled testů a jejich hodnotících vah

6.3. Sestavení vzorce pro výpočet

Samotný vzorec pro výpočet hodnocení musí být připraven na to, že se počet hodnotících prvků může v průběhu času měnit. V tomto směru tedy bude nutné využít počet hodnotících prvků a připravit vzorec univerzálně.

Jak bylo popsáno v předchozí kapitole, každý test má určenou svou váhu. Některé testy mohou mít hodnotící váhu menší než 1 některé naopak větší než 1. Každý test má přiřazeny své koeficienty, kterými se ohodnocuje konkrétní výsledek testované webové stránky. Vzorec pro výpočet je reprezentován celkovým součtem násobků tří prvků:

- 1) hodnotící váhy testu
- 2) hodnotícího koeficientu testu
- 3) ohodnocení na test

Jak již bylo uvedeno, prvky 1) a 2) byly předem stanoveny. Ohodnocení na test je dáno podílem maximálního ohodnocení (10000 bodů) a celkovou hodnotící vahou, která je dána součtem vah všech testů. Samotný vzorec pro představu můžeme popsat následujícím zápisem v metajazyce:

Maximalni_hodnoceni = 10000;

Celkova_hodnotici_vaha = sum(vaha_vsech_prvku);

Hodnoceni_na_prvek = maximalni_hodnoceni / celkova_hodnotici_vaha;

Vysledek = sum(hodnotici_vaha_prvku * hodnotici_koeficient_prvku * hodnoceni_na_prvek);

Zaokrouhleny_vysledek = zaokrouhli_na_cela_cisla (vysledek/100);

6.4. Nastavení vzorce

Statistická analýza, která je popsána v kapitole 8 jednoznačně neprokázala závislost parametrů na pozici ve vyhledávači. Proto bylo finální nastavení vzorce, zejména koeficientů, nastaveno dle subjektivního pohledu autora. Subjektivní pohled byl tvořen empirickým zkoumáním grafů jednotlivých parametrů z výsledků *Trend testů*. Vybrané koeficienty jsou zobrazeny v tabulce 5.

Přehled koeficientů			
Typ	Hodnota 1	Hodnota 2	Koeficient
Google pozice			
od - do	1	10	1
od - do	11	20	0.9
od - do	21	30	0.8
od - do	31	50	0.6
od - do	51	100	0.4
od - do	101	250	0.2
Větší než	250	---	0
Seznam			
od - do	1	10	1
od - do	11	20	0.9
od - do	21	30	0.8
od - do	31	50	0.6
od - do	51	100	0.4
od - do	101	250	0.2
Větší než	250	---	0
Google Pagerank			
Je rovno	10	---	1
Je rovno	9	---	0.9
Je rovno	8	---	0.8
Je rovno	7	---	0.7
Je rovno	6	---	0.6
Je rovno	5	---	0.5
Je rovno	4	---	0.4
Je rovno	3	---	0.3
Je rovno	2	---	0.2
Je rovno	1	---	0.1
Je rovno	0	---	0

Tabulka 5: Přehled vybraných koeficientů

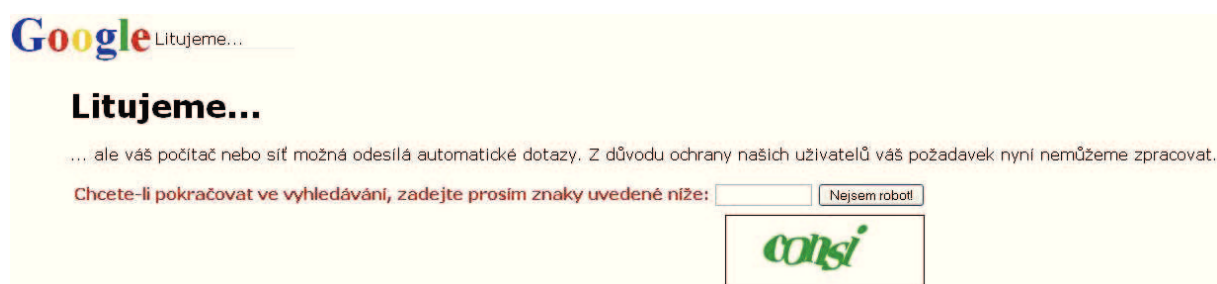
7. Problémy při realizaci

7.1. Ochrana proti robotům

Tento projekt využívá informací dostupných na jiných serverech, které se hromadnému načítání informací pomocí robotů brání. Jedním z problémů tedy bylo řešení, které by mohlo tuto ochranu obejít. Vyhledávače Seznam a Google obvykle zaznamenají zvýšený počet dotazů v krátkém časovém úseku. A následně zobrazí ověřovací formulář, ve kterém musíme opsat znaky z obrázku, a tím se vyhledávače ujistí, že nejsme roboti.

Tento problém byl vyřešen tak, že byly vytvořeny pomocné nástroje, které se neustále dotazují na jistou první pozici ve vyhledávání. Např. pomocí klíčového slova „sport“ najdeme na první pozici *www.sport.cz* (nutno předem ověřit). Zadáme tedy do nástroje v administraci tato kritéria a časovou prodlevu, o které předpokládáme, že bude dostačující. Např. 3s. Nástroj v intervalu 3s posílá dotazy na vyhledávač a zastaví se v případě, že výsledek není dostupný. Jak již bylo zmíněno v předchozím odstavci, vyhledávač nás odmítne obsloužit a zobrazí nám formulář s obrázkem a tak již není zadaná URL adresa nalezena. Cca po 30min zvýšíme prodlevu na 4s a znovu testujeme. Tyto kroky opakujeme tak dlouho, dokud nám nástroj nevrátí dostatečný počet výsledků.

Takto jsme získali časový interval mezi jednotlivými testy týkající se jednoho vyhledávače. Tímto nastavením se sice prodlužuje doba zpracování celkového testu a snižuje se kapacita zkontrolovaných URL adres, ale zajistíme tak kontinuální běh robota.



Obrázek 15: Omezení přístupu robotům Google.com

7.2. Relevantnost výsledků

Dalším podstatným problémem se ukázalo ověření, zda testy vracejí správné výsledky. Tento problém měl pouze jedno řešení. Snaha co nejlépe navrhnout daný test a ruční ověřování tisíců výsledků. Jednotlivé výsledky, které testy poskytly, byly ověřovány v dalších nástrojích nebo přímo na serverech, ze kterých se dané informace získávají. Např. validita stránky byla testována přímo pomocí validátoru na stránkách poskytovatele této služby a srovnány s výsledky, které vrátil test. Toto ověřování se nakonec ukázalo jako časově velmi náročný proces.

Testování práce robota, ověření relevantních výsledků i po hodinách práce a doladění funkčnosti robota si vyžádalo téměř 1000 spuštění. Jelikož jsem dbal na pravdivost výsledků, v případě zjištění i částečné nefunkčnosti byly smazány stovky již hotových výsledků.

7.3. Google API

Vyhledávač Google poskytuje tvůrcům webových projektů nástroje, kterými je možné využívat služeb Google a které mohou implementovat pomocí rozhraní API do svých projektů. Jedním z takto poskytovaných API je i Google Web Search. Toto rozhraní umožňuje posílat dotazy na servery Googlu, které nahrazují dotazy na výsledky vyhledávání pomocí prohlížeče na serveru *www.google.com*. Server pak vrací ve strukturované podobě výsledky na klíčový výraz, které je možné dále zpracovat.

Nevýhodou tohoto API je, že při posílání požadavku je nutno zadat tzv. API KEY, což je klíč, který je přiřazen každému registrovanému uživateli. Bez tohoto API KEY není možné službu využívat. K jednotlivému API KEY je pak přiřazen limit 100 dotazů za den. Je tedy možné poslat pouze 100 požadavků na hledání výsledků. Tento limit by byl překročen během několika minut, protože robot testuje cca 1000 URL adres za den a při každém testování se provede několik dotazů na servery Googlu. Proto bylo nutné dotazy na servery Googlu implementovat tak, aby simulovaly dotazy přímo z prohlížeče.

7.4. Spolehlivá práce robota

Některé testy vyžadují více času na zpracování a jsou prováděny v iteracích o předem neznámém počtu opakování. Jedná se zejména o mapování, které může trvat významně dlouhou dobu. Výhodou se pak ukázalo použití Frameworku JQuery, který pomocí AJAX¹¹ volání umožňuje spouštět funkce v neomezeném počtu iterací a zpracovávat výsledek postupně. Tento způsob volání funkcí pak umožnil vytvořit aplikaci robota, běžící v internetovém okně prohlížeče.

¹¹ AJAX - je obecné označení pro technologie vývoje interaktivních webových aplikací, které mění obsah svých stránek bez nutnosti jejich znovunačítání. Na rozdíl od klasických webových aplikací poskytují uživatelsky příjemnější prostředí, ale vyžadují použití moderních webových prohlížečů [3]

8. Statistické vyhodnocení

V běžném životě určujeme pořadí podle určitých parametrů. Při závodech běžců je pořadí určeno dosažením nejlepšího (nejmenšího) času. U některých znalostních soutěží je vítěz ten, kdo získá nejvíce bodů nebo udělá méně chyb. Vyhledávače a hodnotící algoritmy postupují podobným způsobem. Hodnotící algoritmy však vyhodnocují mnohem větší počet hodnot než jen dosažený čas. V této práci se však zaměříme pouze na závislosti jednotlivých parametrů na pozici ve výsledcích vyhledávání.

Využijeme tento předpoklad a pokusíme se ověřit, zda námi testované parametry jednoznačně ovlivňují pozici ve vyhledávači. Pozorovat budeme hodnoty získané pomocí *Trend testů*. Ke zpracování těchto hodnot použijeme statistický software Statgraphics. Použitou metodou bude *Simple Regression*, jejím výsledkem je tabulka analýzy rozptylu.

Jednou z početních statistických metod jak ověřit závislost dvou prvků je analýza rozptylu. Jde o metodu matematické statistiky, která umožňuje ověřit, zda na hodnotu náhodné veličiny pro určitého jedince má statisticky významný vliv hodnota některého znaku, který se u jedince dá pozorovat. Analýza rozptylu je pro víc než jeden znak značně výpočetně náročná metoda a je pro ni téměř vždy potřeba počítač se speciálním statistickým softwarem. Pro interpretaci tabulky jsou nejdůležitější *p hodnoty*. [4]

Pro zjednodušení si postačíme s faktem, že je-li *p hodnota* pro dané hodnoty vkládané do analýzy rozptylu menší než 0.05 tak lze prohlásit, že mezi nimi existuje statisticky významná závislost. Kompletní statistická data, která byla zpracována, jsou dostupná v administraci v sekci *Trendy – Výsledky trendů* - po kliknutí na ikonku tabulky. Hodnotami vstupujícími do testu jsou vždy pozice ve vyhledávači (Seznam nebo Google) a hodnoty výsledků jednotlivých testů. Výsledky všech testů jsou uloženy v tabulce na příloženém CD.

Parametry, které nejčastěji vykazovaly statisticky významnou závislost, ukazuje následující tabulka 6. Červeně vyznačené hodnoty pak potvrzují statisticky významnou závislost.

Měření	Vyhledávač	Klíčové slovo	Google Rank	Seznam Rank	Index Google	Alexa Rank	Zpětné odkazy
Únor 2011	Google	Studium	0,49	0,0004	0,1937	0,044	0,0001
Únor 2011	Google	Sport	0,0749	0,5571	0,2299	0,0439	0,0003
Únor 2011	Google	Reality	0,0014	0,0287	0,0023	0	0,0039
Únor 2011	Seznam	Studium	0,3648	0,026	0,3039	0,2663	0,5127
Únor 2011	Seznam	Sport	0,0457	0,0032	0,0219	0,0115	0,0096
Únor 2011	Seznam	Reality	0	0	0,504	0,0007	0,1088
Březen 2011	Google	Studium	0,171	0,0911	0,8361	0,1135	0,0106
Březen 2011	Google	Sport	0,0104	0,0762	0,0001	0,0004	0,0033
Březen 2011	Google	Reality	0,0037	0,0027	0	0,0005	0,0006
Březen 2011	Google	Obrazy	0,3135	0,0026	0,5313	0,3901	0,1326
Březen 2011	Seznam	Studium	0,0037	0,0001	0,0307	0,2592	0,2102
Březen 2011	Seznam	Sport	0,0684	0,0942	0,0048	0,0252	0,0074
Březen 2011	Seznam	Reality	0,0127	0,0318	0,0001	0,0019	0,0314
Březen 2011	Seznam	Obrazy	0,0369	0,3788	0,1129	0,3612	0,2118
Duben 2011	Google	Studium	0,3421	0,0028	0,1567	0,3571	0,0095

Duben 2011	Google	Sport	0,0144	0,1038	0	0	0,0057
Duben 2011	Google	Reality	0,0271	0,0015	0,2484	0,0003	0,0038
Duben 2011	Google	Obrazy	0,0017	0,0014	0,7843	0,49	0,3721
Duben 2011	Seznam	Studium	0,1631	0,0024	0,1425	0,4616	0,1425
Duben 2011	Seznam	Sport	0,3299	0,0572	0,0142	0,2211	0,0069
Duben 2011	Seznam	Reality	0,1679	0,0001	0,4286	0,0083	0,0352
Duben 2011	Seznam	Obrazy	0,9727	0,0394	0,3672	0,1032	0,3983

Tabulka 6: Část výsledků statistických testů

Přestože některé parametry vykázaly závislost, ani u jednoho se neprokázala ve všech testovaných případech, z čehož nelze usoudit, že je tato závislost pravidlem. Nelze tedy jednoznačně prokázat dosažené závislosti a musíme prohlásit, že mezi testovanými parametry a pozicemi ve vyhledávacích Google a Seznam neexistuje statistická závislost. Z tabulky pak můžeme pozorovat, že v měsících Únor 2011 a Březen 2011 se u vyhledávače Google na klíčový výraz *Reality* projevila závislost u všech zmíněných parametrů. V měsíci Duben 2011 se však již tato skutečnost neprojevila. Výsledky statistických testů pouze naznačují, které parametry dosahují nejčastěji závislosti.

Další metodou, kterou využijeme je empirické zkoumání hodnot vynesných do grafu. Jednotlivé výsledky či jejich kombinace zobrazíme v grafu a pozorujeme, zda křivka vykazuje určité chování.

Pokud v grafu bude viditelná alespoň přibližná křivka, která bude klesající nebo rostoucí vzhledem k pozici, pak můžeme o tomto parametru prohlásit, že ovlivňuje pozici URL adresy ve vyhledávači, resp. pozice je závislá na hodnotě tohoto parametru.

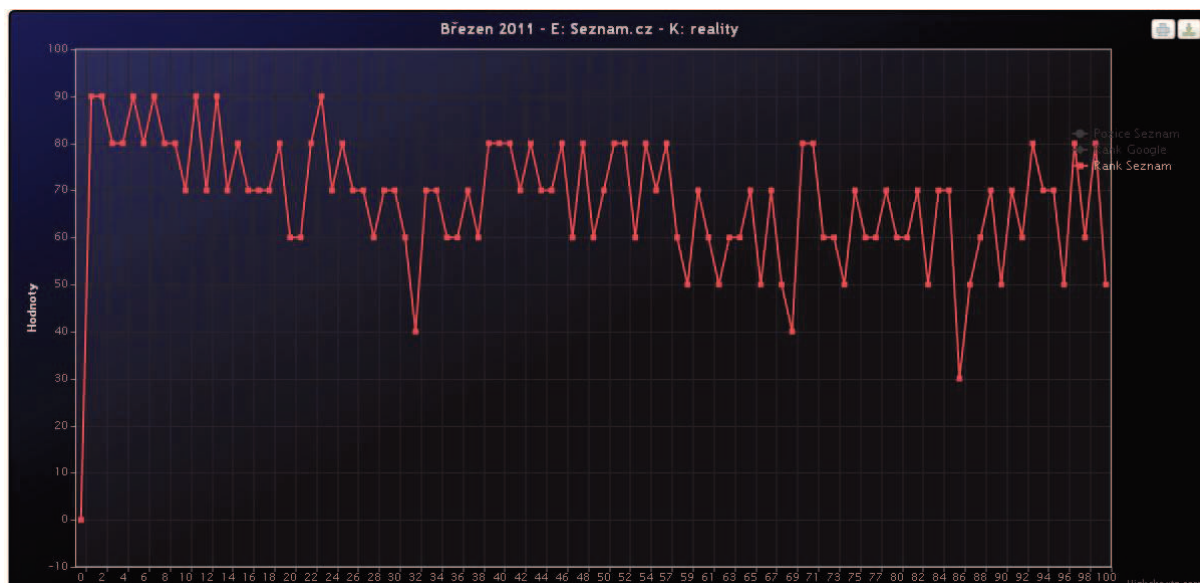
V případě, že se tato závislost nepotvrdí, můžeme tento parametr i dále zkoumat. Můžeme zjistit, zda prvních 100 výsledků dosahuje určitých minimálních hodnot a porovnat parametry mezi sebou a tím ověřit další souvislosti. Z tohoto zkoumání pak můžeme upřesnit koeficienty hodnotícího algoritmu.

V první fázi byly empiricky zkoumány jednotlivé výsledky testů popsané v kapitole 3. Podrobněji jsou popsány pouze ty parametry, které vykázaly určitý typ chování. O parametrech, u kterých se nepodařilo prokázat závislost, se v tomto textu zmiňovat nebudeme.

Data pocházejí z měření z měsíce Únor, Březen a Duben 2011. V měsíci Únor 2011 bylo pro získávání výsledků použito těchto *Trend slov*: *studium*, *sport*, *reality*. Měření v měsících Březen a Duben 2011 byla tato slova rozšířena o slovo "obrazy".

8.1. S-Rank

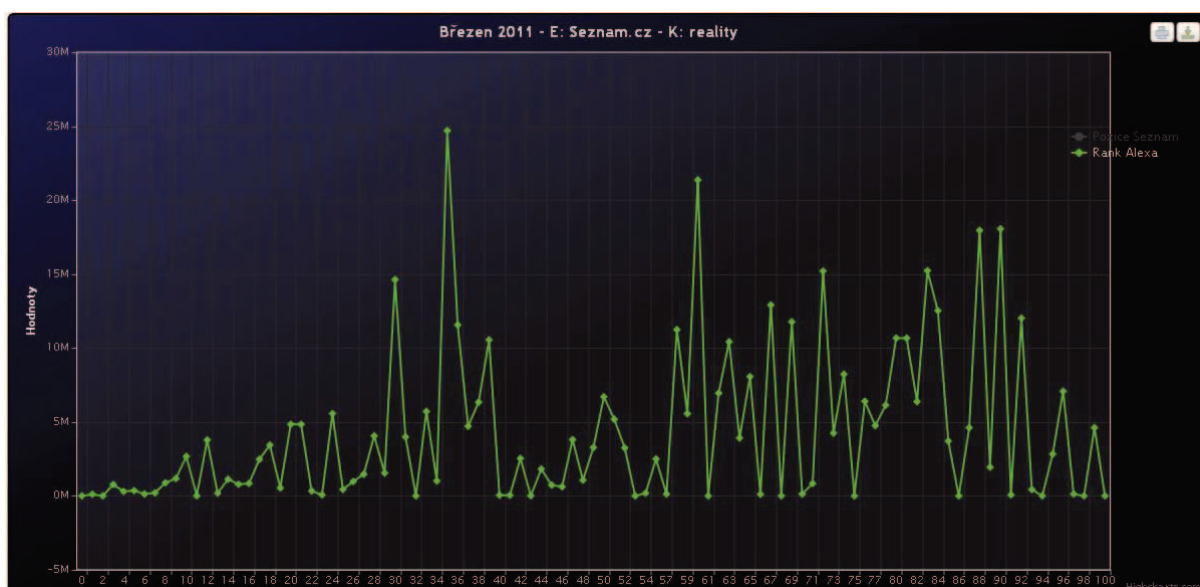
Z výsledků, které byly získány z *Trend testů* provedených na vyhledávači Seznam.cz vyplývá, že URL adresy, které se vyskytují na prvních stranách výsledků hledání, mají vyšší S-Rank než URL adresy, které se vyskytují na horších pozicích. Viz. Obrázek 16.



Obrázek 16: Seznam.cz: Graf S-Ranku v závislosti na pozicích ve výsledcích hledání

8.2. Alexa Rank

Jak již bylo zmíněno v kapitole 3.12, Alexa Rank hodnotí návštěvnost na webových stránkách. Tento Rank dosahuje hodnot od 1 do několika milionů. Z výsledků pak nelze odvodit závislost pozice na hodnotě Alexa Ranku, protože hodnoty nemají ani rostoucí ani klesající tendenci. Stránka na pozici 100 má hodnotu Alexa Ranku pod 5 mil. stejně jako stránky na předních pozicích. Z obrázku 17 je však patrné, že nejvyšší návštěvnost (uživatelů, kteří mají nainstalován Alexa Toolbar) je prvních 20 výsledků. Tedy uživatelé navštěvují URL adresy na prvních dvou stranách výsledků.

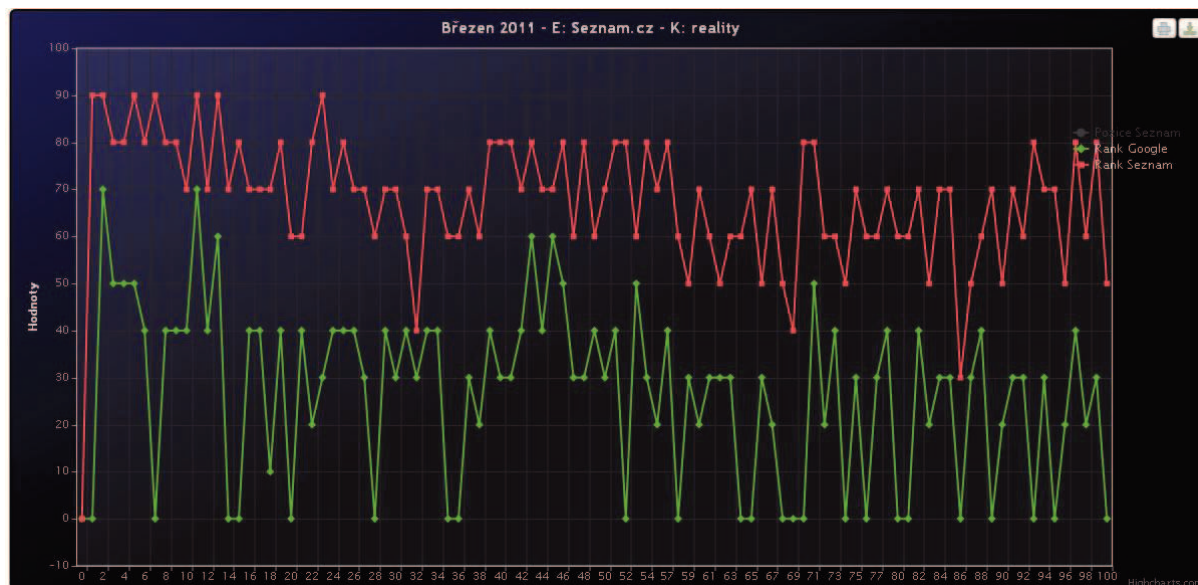


Obrázek 17: Hodnota Alexa Ranku je výrazně nižší u prvních 20 výsledků.

8.3. Google Rank vs. S-Rank

V druhé fázi byly zkoumány kombinace výsledků testovaných parametrů. Z výsledků zkoumání opět zmíníme pouze zjištěné vlastnosti.

Toto srovnání jednoznačně prokázalo, že S-Rank URL adresy je ve většině případů vyšší než Google Rank. Tato vlastnost se na různá *Trend slova* liší. Srovnání je pak zobrazeno v grafu na obrázku 18 (Google Rank je vyznačen v grafu zeleně, Seznam Rank červeně).

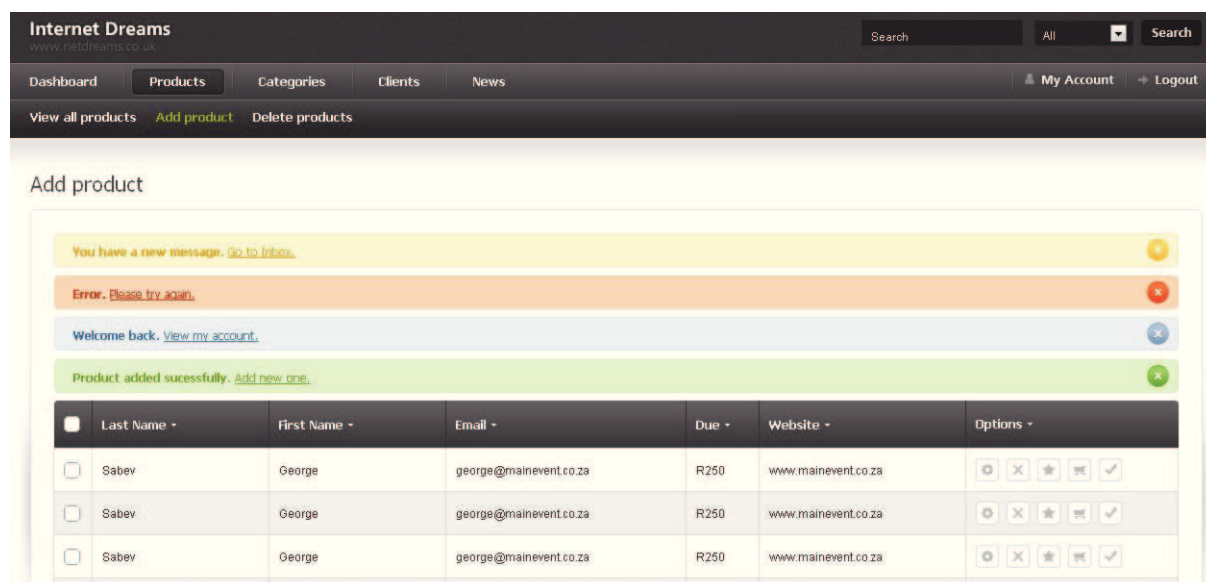


Obrázek 18: Srovnání hodnot S-Rank a Google Rank.

9. GUI

9.1. Grafický layout

Při výběru celkového grafického rozhraní projektu bylo dbáno na to, aby byla grafická stránka projektu příjemná, přehledná a přitom intuitivní pro uživatele. Proto bylo využito bezplatné šablony tvůrce *Internet Dreams Ltd.* www.netdreams.co.uk, která maximálně vyhovovala požadavkům.



Obrázek 19: Vzhled webové aplikace

9.2. Registrace a přihlášení

Pro registraci si vystačíme se základními informacemi, jako je email, uživatelské jméno a heslo. Další informace může uživatel vyplnit a nastavit až po přihlášení. Pro dokončení registrace musí uživatel kliknout na ověřovací odkaz v emailu zasláném na email, který uživatel uvedl při vyplňování registračního formuláře. Pro testování byl vytvořen uživatel *backdoor* s heslem *backdoor*.

The image shows two side-by-side screenshots of web forms. The left screenshot is titled 'Registrace do systému' (Registration to the system) and contains a 'Vytvoření účtu' (Create account) section with fields for 'Email', 'Login', 'Heslo' (Password), and 'Heslo znovu' (Password again), followed by a 'Registrovat' (Register) button. The right screenshot is titled 'Přihlášení do systému' (Login to the system) and contains a 'Zadejte své přihlašovací údaje' (Enter your login details) section with fields for 'Login' and 'Heslo', followed by a 'Login' button. Both forms have a header with 'SEO ROOT' and 'Autor: Tomáš Gottvald'.

Obrázky 20 a 21: Registrace a Přihlášení






9.3. Projekty

Každý uživatel si po registraci může přidat neomezený počet projektů. U projektu zadá URL adresu projektu a klíčová slova, na která chce svůj projekt testovat. Po vytvoření projektu bude uživatel vyzván k umístění ověřovacího souboru do kořenového adresáře webu. Tím bude zajištěno, že projekt bude vkládat buď vlastník anebo osoba pověřená.

Z výpisu uložených projektů je pak patrné, které projekty jsou ověřeny, jaká je jich zadaná URL adresa, datum posledního načtení projektu robotem. Dále pak jsou k dispozici nástroje na editaci údajů o projektu, statistiky projektu - výsledky získané testováním robota a další doplňující nástroje pro tvůrce stránek nebo majitele projektu.

Výpis projektů

Máte vytvořeny tyto projekty

Ověřeno	URL projektu	Keywords	Poslední kontrola	Nástroje
	http://www.neklíkejte.cz	tvorba www	24.04.2011	  
	http://www.obrazyshop.cz	keramické obrazy	24.04.2011	  

Obrázek 22: Výpis projektů

9.4. Statistiky

Po kliknutí na tlačítko *Statistiky* se uživateli zobrazí jednotlivé stránky, které robot našel při procházení stránek projektu a jsou označeny jako *Kontrolované stránky*. Stránky, které robot z různých důvodů ignoruje, jsou v tabulce *Ignorované stránky*. Tabulka je v případě velkého počtu URL adres stránkována po 100 řádcích.

K jednotlivým URL adresám se uživateli zobrazí poslední dosažené výsledky. Protože mají výsledky všech testů v tabulce nadměrnou šířku a není možné je na standardním rozlišení zobrazit všechny najednou, má uživatel k dispozici filtr výsledků. Pomocí tohoto filtru si může vybrat výsledky, které chce zobrazit.

Může si také k jednotlivým projektům uložit i vlastní filtr. V editaci projektu, si vybere pomocí zaškrtačích políček, které testy chce u daného projektu zobrazovat a uloží změny. Při dalším prohlížení statistik se již automaticky nastaví takto uložený filtr. Filtr také slouží k lepší orientaci ve výsledcích a je určen pro uživatele, kterým postačí pouze některé informace, které považuje za důležité.

Filtr:

Test	Pozice Google	Pozice Seznam	Google Rank	Seznam Rank	Index Google	Index Seznam	Alexa Rank	Zpětné odkazy	Popularita	Seznam: rozšířená shoda	Seznam: přesná shoda
Zobrazit	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

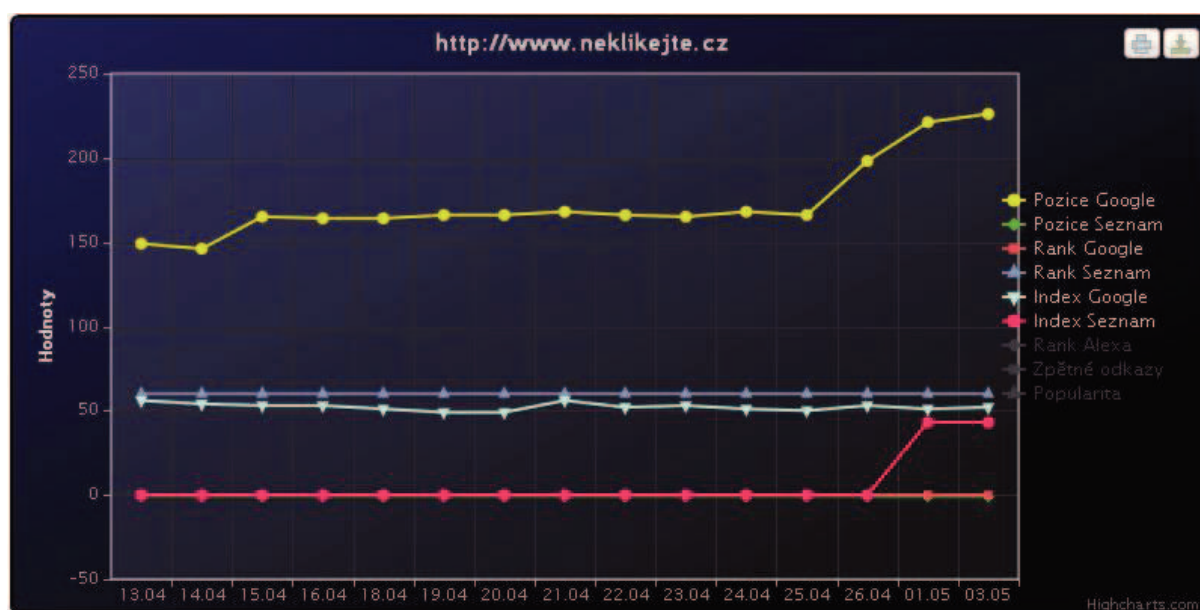
Změnit filtr

Obrázek 23: Ukázka možnosti nastavení filtru

Tabulka kontrolovaných stránek pak zobrazuje URL adresy, které jsou zařazeny do fronty testování robotem. Nad URL adresami jsou 3 symboly. Po kliknutí na první reprezentující zeměkouli se stránka zobrazí v novém okně prohlížeče.

Druhý symbol, žlutá hvězdička, obecně známý z internetových prohlížečů, pak nabízí uživateli zařadit URL adresu do tabulky *Oblíbené stránky*. Tuto funkci pak mohou využít uživatelé, kteří mají přidány projekty, obsahující velký počet URL adres. Kliknutím na tento symbol hvězdička zmizí a při obnovení stránky se bude tato URL adresa zobrazovat v tabulce *Oblíbených stránek*.

Kliknutím na třetí symbol může uživatel URL adresu a výsledky k ní dosažené ze systému smazat. Pro jednotlivé URL adresy je zobrazeno na konci řádku tabulky tlačítko *Graf*. Po kliknutí na něj se zobrazí historický vývoj hodnot, které si uživatel zvolil pomocí filtru. Pod grafem pak může uživatel filtrovat výsledky dle data.





Obrázek 24: Statistika URL adresy zadané uživatelem

Tabulka Ignorovaných stránek pak uživateli umožňuje URL adresu smazat, pokud ví, že se již na stránkách projektu nenachází – např. přejmenuje-li kontakt.php na kontakty.php. Pokud je URL adresa ignorována, protože nebyla tři testování po sobě dostupná tak ji uživatel může vrátit zpět do testování. Po kliknutí na tlačítko ve sloupci *Neignorovat* je proveden test dostupnosti, zároveň se ověří, zda URL adresa směřuje na doménu projektu. Posledním testem, který proběhne, než se stránka vrátí, zpět mezi *Kontrolované stránky* je test, zda URL adresa neobsahuje nepovolenou příponu – jak bylo popsáno v kapitole 5. Pokud všechny testy proběhnou v pořádku stránka je smazána z ignorovaných a je zařazena do testování.

9.5. Nástroje pro tvůrce stránek

Nástroje pro tvůrce stránek využívají některých vlastností testů, které využívají informací (zdrojového kódu) dostupných přímo na stránkách projektu. Vzhledem k problematice Ochrany proti robotům tedy nebylo možné využít testů, které získávají informace z vyhledávačů, protože by byl v krátkém čase překročen limit počtu dotazů v určitém časovém úseku daný jednotlivými vyhledávači.

Nástroje	
Test	Spuštění testu
Mapování a sitemapa stránky	
Test validity a dostupnosti	
Jak vidí robot Vaši stránku (prohlížeč LYNX)	

Obrázek 25: Výpis doplňujících nástrojů pro uživatele

9.5.1. Mapování a sitemapa stránky

Po přidání projektu je do testovaných stránek vložena URL adresa projektu. Od té doby robot automaticky mapuje jednotlivé stránky projektu a doplňuje nově nalezené URL adresy. Pokud uživatel nechce čekat, až robot vytvoří sitemapu stránky automaticky, má možnost spustit tento nástroj ručně.

Po spuštění testu jsou testovány jednotlivé dostupné stránky, uživatel vidí procentuální vyjádření dokončení testu. Po dokončení testu vypíše výsledek testu, tedy počet nově nalezených URL adres. Ty jsou automaticky doplněny do seznamu testovaných stránek.

Uživatel má možnost spustit celý test znovu, nebo označit stránky, které chce testovat znovu nebo spustit test pouze nově nalezených stránek. Tímto nástrojem si může ověřit dostupnost jednotlivých podstránek. Nástroj smí používat neomezeně. Po několikerém spuštění si může uživatel jedním kliknutím vygenerovat soubor sitemap.xml, který pak použije pro svůj projekt.

9.5.2. Test validity a dostupnosti stránek

Pokud uživatel dbá na validitu stránek, má zde k dispozici účinný nástroj, který vyhodnotí validitu všech dostupných stránek v jednom kroku. Odpadá tak složité proklikávání celého obsahu projektu a zjišťování validity pro jednotlivé stránky. Při spuštění tohoto nástroje tak dostává uživatel informace, zda je stránka, získaná pomocí mapování, dostupná, jakou má velikost a zda je validní. Pokud validní není, je uživateli zobrazen odkaz, který vede na nástroj W3 validátoru, který popisuje jednotlivé chyby.

Po ukončení testování má uživatel možnost spustit test vybraných položek, spustit test pouze na konkrétní URL adresu nebo nechat testovat pouze nevalidní stránky. Takto si může průběžně kontrolovat validitu svého projektu.

Stejně jako u předchozího nástroje, je upozorněn vyskakovacím oknem ihned po ukončení testování. Uživateli jsou během testování zobrazovány výsledky pro jednotlivé testované stránky. Nemá tak pocit, že se nic neděje (jako je běžné u jiných nástrojů).

9.5.3. Zobrazení Lynx

Existuje mnoho nástrojů, které zobrazí stránku tak jak ji s největší pravděpodobností "vidí" vyhledávač. Ten přistupuje ke stránce jako k textu, který uloží a dále analyzuje a ohodnocuje. K tomuto účelu byl použit nástroj Lynx, který je dostupný na URL adrese:

<http://crschmidt.net/services/lynx/> Za poslední lomítko pak nástroj přidá URL adresu vybrané stránky bez počátečního "http://" a "www. " a výsledek zobrazí uživateli v novém okně prohlížeče.

```
[1]neklikejte Vase ID _____
Přihlásit
* [2]AKCE
* [3]HOME
* [4]NASE SLUZBY
* [5]CHCI WEB
* [6]JAK PRACUJEME
* [7]REFERENCE
* [8]KONTAKTY

hotline HOTLINE kontakt: 773 834 134, info@neklikejte.cz
* [9]Eshopy
* [10]Webové prezentace
* [11]PC servis, IT servis
* [12]SEO optimalizace
* [13]Redakční systémy
* [14]Informační systémy
* [15]Lokální aplikace
* [16]Webové aplikace

neklikejte.cz
```

Obrázek 26: Ukázka výpisu z prohlížeče Lynx

9.6. Diskuze

Pro registrované uživatele je připravena sekce *Diskuze*, která je velmi podobná zdi na Facebooku. Tento model byl vybrán vzhledem k tomu, že je hojně využíván a uživatelé jsou na něj zvyklí. Uživatelé mohou vytvářet libovolná témata diskuzí a konzultovat tak svoje postřehy nebo názory.

9.7. Administrace

Administrace projektu je součástí webové části projektu ve složce "admin". Pro vstup do administrace je nutné zadat za URL adresu, kde je projekt umístěn, slovo admin ("...URL projektu.../admin/"). Administrace nabízí možnosti spuštění robotů, přípravu testů, definování Trend slov a obsahuje i další nástroje pro administraci projektu. Aplikace je dočasně dostupná na URL adrese: <http://www.neklikejte.cz/develop/seo/>, i do administrace je možné se přihlásit pod uživatelským jménem *backdoor* a heslem *backdoor*.

9.7.1. Bots

Sekce *Bots* obsahuje obsluhu robotů, kteří automaticky testují jednotlivé URL adresy. Běžný robot testuje jednotlivé URL adresy registrovaných projektů. Pokud není robot spuštěn přímo na serveru, lze jeho spuštění provést stisknutím tlačítka *Start*. Jeho činnost lze zastavit kliknutím na tlačítko *Stop*. Pro správnou a nepřetržitou funkci robota je nutno ponechat okno webového prohlížeče otevřené právě v sekci *Běžný robot*. Robot průběžně zobrazuje dosažené výsledky a tak je administrátor informován, která stránka je testována, mapována, a které údaje byly naposledy uloženy. Trend bot testuje stránky připravených Trend Testů. Jeho vizuální podoba i obsluha je totožná s obsluhou běžného robota.

9.7.2. Trendy

Tato sekce umožňuje připravovat a nastavovat testy pro *Trend bota*. Pro kompletní přípravu testů je nutné nastavit *Trend slova*. Trend slova jsou slova, která budou použity při získání prvních 100 URL adres z výsledků vyhledávačů. Na tyto URL pak bude *Trend bot* provádět testy. Výsledky těchto měření slouží jako podklady pro statistické vyhodnocení. *Trend slova* lze v této sekci přidávat a mazat, nelze je editovat, aby nebyly ovlivněny již dosažené výsledky.

V sekci *Připravit test* je možné nadefinovat parametry *Trend Testu*. Je nutné zadat název testu, aby bylo možné výsledky oddělit. Administrátor může vybrat pouze některá vložená *Trend slova* a pouze některé vyhledávače nebo může vybrat všechna slova a všechny vyhledávače. Na počátku přípravy testu je vybráno vše. Po kliknutí na tlačítko "Vytvořit test" se na vybraná *Trend slova* provede dotaz na zvolený vyhledávač a uloží prvních 100 URL adres vrácených vyhledávačem. Tento krok se provede pro všechna vybraná slova a pro všechny vybrané vyhledávače. Následně je nutné spustit Trend Bota, který na takto připravené URL adresy provede výše popsané testy.

Dosažené výsledky *Trend testů* jsou pak dostupné v sekci *Výsledky testů*. Tabulka je rozdělena záhlavím tabulky podle jednotlivých názvů testů. Pro přehlednost název testu reprezentuje měsíc a rok, ve kterém bylo testování provedeno. Srovnáním dosažených výsledků pak lze sledovat hodnotící trendy vyhledávačů a jejich vývoj. Každý řádek tabulky pak představuje kombinaci *Trend slova* a vyhledávače, ze kterého výsledky pocházejí. Na konci každého řádku je pak k dispozici nástroj, který zobrazí předpřipravené filtry pro zobrazení výsledků. Tyto filtry lze pak připravit v administraci. Těmito filtry pak lze nastavit např. zobrazení výsledků Google Rank a S-Rank v závislosti na pořadí odkazu. Na konci každého řádku s takto připravenými filtry je tlačítko, které pak zobrazí výsledný graf, ze kterého lze provést pozorování.

9.7.3. Algoritmus

V této sekci administrace je možné ovlivnit chování hodnotícího algoritmu. V sekci *Testy* je možné evidovat jednotlivé testy a jejich hodnotící váhy. V sekci *Parametry testů* pak může administrátor spravovat hodnotící koeficienty a jejich hodnoty pro jednotlivé testy. Jednotlivé koeficienty lze evidovat nebo úplně smazat. Možnost přidání koeficientu se nachází ve spodní části stránky. Při zadávání koeficientu je nutné vyplnit název testu, typ koeficientu, kde má administrátor na výběr z možností *Rovná se (hodnota 1)*, *Menší než (hodnota 1)*, *Větší než (hodnota 1)* nebo *Od hodnoty 1 do hodnoty 2*. Následně pak definuje jednotlivé hodnoty a koeficient, který bude dosazen, pokud bude výstupní hodnota testu splňovat definované požadavky. Lze tak například nastavit, že bude-li velikost stránky od 0 do 30 kb pak se výsledku tohoto testu přiřadí koeficient 1.

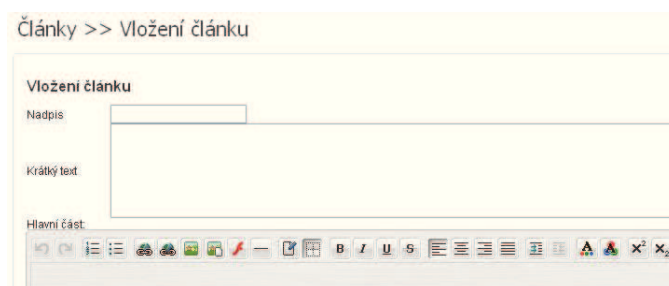
9.7.4. Filtry

Jednou z variant jak posoudit vliv testovaných parametrů na pořadí ve výsledcích vyhledávání se provede pomocí zkoumání křivky dané grafem výsledků. Kdybychom však všechny výsledky chtěli vyznačit v grafu, byl by dosti nepřehledný a měřítko některých hodnot by nám nedovolovalo správné vyhodnocení (např. Alexa Rank dosahuje milionů a google rank hodnot do 10). Tento problém se dá vyřešit vypínáním a zapínáním dostupných hodnot. Tento způsob je však časově a výpočetně náročný a celou analýzu výsledků znepříjemňuje.

Proto si v sekci *Filtry* můžeme předpřipravít kombinace parametrů, které chceme pro každé měření, Trend slovo a vyhledávač používat častěji. Pod tabulkou zobrazující přehled již předpřipravených filtrů si administrátor může nadefinovat filtr nový. Pro větší přehled vyplní jeho název (např. Srovnání S-ranku a G-ranku), popis testů, který by měl stručně charakterizovat smysl testu a zaškrtnutím políček pod názvy jednotlivých testů přiřadí zobrazované hodnoty.

9.7.5. Články

V této sekci může administrátor vkládat a editovat vložené články. Může tak zveřejnit a prezentovat výsledky vyhodnocení *Trend testů* široké veřejnosti, popřípadě upozorňovat na důležité články z oblasti SEO. Články může vkládat pomocí WYSIWYG editoru¹².



Obrázek 27: WYSIWYG editor v administraci pro psaní článků

¹² WYSIWYG editor – z anglického What You See Is What You Get – v překladu To co vidíte, to dostanete. Editor umožňující editovat HTML kód v grafické podobě. Nástroj je velmi podobný známé aplikaci MS Word.

10. Závěr

Cílem této diplomové práce bylo vytvoření SEO analyzátoru, který bude testovat jednotlivé, komunitou uznávané, SEO parametry. Byl vytvořen robot, který automaticky testuje vytvořené projekty, prochází všechny dostupné podstránky projektů a poskytuje uživateli přehlednou formou získané výsledky.

Vlastnosti tohoto robota byly využity k získávání informací, které umožnily pomocí statistických metod ověřit závislosti testovaných parametrů na pozici ve vyhledávačích Google a Seznam. Tímto byly veřejně prezentovány podklady, které pomohly vyvrátit nebo potvrdit určité skutečnosti, o kterých se v komunitě tvůrců webových stránek vedly vášnivé debaty. SEO analyzátor je možné i nadále využívat a v pravidelných intervalech přezkoumávat tyto informace a v relativně krátkém čase tak získávat přehled o trendech v hodnocení stránek vyhledávači.

Při psaní této diplomové práce jsem četl velké množství diskuzí, jejichž příspěvky byly ve velké míře rozporuplné, což mne pouze utvrzovalo ve velkém významu tohoto výzkumu. Jednotlivá tvrzení členů SEO komunity tak již nemusí být založeny na domněnkách nebo částečných informacích, ale nyní je bylo možné podložit fakty a dále nad nimi diskutovat.

Přestože bylo použito programovacího jazyku PHP, bylo evidentní, že použitím pouze tohoto jazyka nebude možné robota efektivně sestrojít. Při práci jsem tak ocenil univerzálnost javascriptového frameworku JQuery, který jsem se naučil ovládat a využívat jeho výhod.

Věřím, že si výsledky tohoto výzkumu najdou mnoho odpůrců, kteří budou použitou techniku zpochybňovat. Je však možné, že tento projekt bude inspirací dalším programátorům. SEO analyzátor neměl ani ve svém počátku ambice změnit logiku hodnocení webů, nesnažil se nahradit současné nástroje, ale je možné, že se tvůrci webových stránek zaměří více na kvalitu textů a obsahu stránek místo toho, aby se umělými způsoby snažili vylepšit testované SEO parametry, protože i podle posledních zpráv od tvůrců vyhledávačů trend spíše k hlubší analýze textů a tím i zvyšování relevantnosti výsledků hledání.

11. Reference

- [1] Michal Kubiček, Jan Linhart: *333 tipů a triků pro SEO: Sbírká nejlepších technik optimalizace webů pro vyhledávače*, vydání první, Computer Press, a.s., 2010, ISBN 978-80-251-2468-0
- [2] Seznam.cz – *Nápověda popisující význam shody klíčových slov*
<<http://napoveda.sklik.cz/cz/vytvarime-kampan/klicova-slova/shody-klicovych-slov/>>
[online] [cit. 2011-01-15]
- [3] Wikipedia.org – *AJAX*
<<http://cs.wikipedia.org/wiki/AJAX>> [online] [cit. 2011-03-02]
- [4] Wikipedia.org – *Analýza rozptylu*
<http://cs.wikipedia.org/wiki/Anal%C3%BDza_rozptylu> [online] [cit. 2011-04-20]
- [5] Jennifer Grappone, Gradiva Couzin: *SEO Search Engine Optimization*, vydání první, ZONER Software, s.r.o., 2007, ISBN 978-80-86815-85-5

PŘÍLOHY

I. Obsah příloženého CD

Obsah CD, přiloženého na poslední straně této diplomové práce:

Složka "PHP/"

Obsahuje všechny zdrojové kódy, potřebné k provozu aplikace.

Složka "DOCS/"

Obsahuje veškerou dostupnou dokumentaci k tomuto projektu.

- složka USER – obsahuje uživatelskou příručku v elektronické podobě
- složka ADMIN – obsahuje administrátorskou příručku v elektronické podobě
- složka PROJECT – obsahuje tento dokument v elektronické podobě
- složka STATS – obsahuje kompletní výsledky statistického zkoumání v elektronické podobě

Složka "MySQL/"

Obsahuje dump soubor, ve kterém jsou definovány všechny tabulky a základní data celé databáze